

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**



ПОБЕДИТЕЛЬ КОНКУРСА ИННОВАЦИОННЫХ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ ВУЗОВ

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК

Выпуск 51

**НАУЧНЫЕ ШКОЛЫ
В СПбГУ ИТМО**



**САНКТ-ПЕТЕРБУРГ
2008**

В научно-техническом вестнике СПбГУ ИТМО, Выпуск 51 «НАУЧНЫЕ ШКОЛЫ В СПбГУ ИТМО» представлены работы, выполненные в рамках:

- инновационной образовательной программы «Инновационная система подготовки специалистов нового поколения в области информационных и оптических технологий» вузов России на 2007–2008 гг.;
- аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы (2006–2008 гг.)» (Федеральное агентство по образованию);
- Федеральной целевой программы развития образования на 2006–2010 гг. (Федеральное агентство по образованию);
- Федеральной целевой программы развития научно-технологического комплекса России на 2007–2012 гг. (Федеральное агентство по науке и инновациям);
- Российского фонда фундаментальных исследований, а также инициативные разработки.



В 2007 году СПбГУ ИТМО стал победителем конкурса инновационных образовательных программ вузов России на 2007–2008 годы. Реализация инновационной образовательной программы «Инновационная система подготовки специалистов нового поколения в области информационных и оптических технологий» позволит выйти на качественно новый уровень подготовки выпускников и удовлетворить возрастающий спрос на специалистов в информационной, оптической и других высокотехнологичных отраслях экономики.

ISSN 1819-222X

© Санкт-Петербургский государственный университет
информационных технологий, механики и оптики, 2008

**Научная школа
«ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ,
ПРОЕКТИРОВАНИЕ,
ТЕХНОЛОГИИ ЭЛЕМЕНТОВ И УЗЛОВ
КОМПЬЮТЕРНЫХ СИСТЕМ»**

**ПРИМЕНЕНИЕ ЭЛЕМЕНТОВ ТЕОРИИ ГРАФОВ
В КОНЕЧНО-ЭЛЕМЕНТНОМ АНАЛИЗЕ****Д.А. Боголюбов****Научный руководитель – к.т.н., доцент Н.С. Кармановский**

Статья подготовлена в рамках проведения инженерных расчетов конструктивов, разрабатываемых в Российском институте радионавигации и времени для глобальной навигационной спутниковой системы ГЛОНАСС. Рассматриваются основы методов решения уравнений физики в механических системах автоматизированного проектирования с помощью методов конечных элементов, излагаются достоинства и недостатки метода, предлагаются возможные способы ликвидации недостатков путем создания принципиально новых алгоритмов, основанных на векторном методе конечных элементов.

Общие теоретические сведения

Метод конечных элементов (МКЭ) – один из самых известных методов решения уравнения физики, реализованных в механических САПР. В настоящее время он является стандартом при решении задач механики твердого тела посредством численных алгоритмов. В основе метода лежит дискретизация объекта с целью решения уравнений механики сплошной среды в предположении, что эти соотношения выполняются в пределах каждой из элементарных областей. Эти области называются конечными элементами. В пределах конечного элемента назначаются свойства ограничиваемого им участка объекта и описываются поля интересующих величин. Параметры из второй группы назначаются в узлах элемента, а затем вводятся интерполирующие функции, посредством которых соответствующие значения можно вычислить в любой точке внутри элемента или на его границе. Задача математического описания элемента сводится к тому, чтобы связать действующие в узлах факторы. Имея математический аппарат для получения матриц жесткости конечных элементов, приведения нагрузок, приложенных к поверхности или в объеме элемента к усилиям в узлах, а также решения обратных задач: вычисления полей деформаций и напряжений в объеме элемента на базе перемещений в узлах, можно построить алгоритм МКЭ.

Использование метода конечных элементов требует решения системы уравнений равновесия вида

$$Ku=R, \tag{1}$$

где u и R – соответственно векторы обобщенных перемещений и нагрузок, K – глобальная матрица жесткости, обычно симметричная, положительно определенная и разреженная.

Теория графов и метод конечных элементов

Применение теории графов при разложении матрицы жесткости широко освещалось в зарубежной технической литературе ([2, 3]). Что же касается решения проблем, возникающих при моделировании процессов, описываемых векторными переменными (в частности, задачи газо- и гидродинамики), до сих пор не существует единообразного подхода к их решению.

В 1980 г. в статье Ж.С. Неделека [6] были представлены новые семейства неконформных конечных элементов – конечные элементы на кубах и тетраэдрах. Эти семейства являются конформными в пространствах $H(\text{rot}, \Omega)$ и $H(\text{div}, \Omega)$, что позволило предложить использование этих элементов для аппроксимации уравнений Максвелла и

некоторых других [4]. Сейчас такие элементы в том или ином виде используются во многих инженерных системах [1].

В настоящее время существует два основных направления развития конечно-элементного анализа: узловый и векторный. Несмотря на имеющиеся различия, узловый и векторный методы конечных элементов имеют общую идеологию. В частности, эта общая идеология проявляется в методологии решения краевых задач методом конечных элементов (как скалярным, так и векторным). Оно состоит из следующих этапов:

1. переход от исходной постановки к эквивалентной вариационной;
2. дискретизация области решения (т.е. ее разбиение на геометрически конечные элементы);
3. выбор интерполяционных функций, определяющих свойства решения;
4. генерация системы линейных алгебраических уравнений;
5. решение сгенерированной системы уравнений [4].

Несмотря на то, что общие принципы векторного и узлового методов конечных элементов достаточно схожи, между ними имеется существенное геометрическое различие: в узловом МКЭ неизвестные величины ассоциированы с вершинами (т.е. точками дискретизации области), а в векторном – с ребрами, т.е. отрезками, соединяющими точки дискретизации. Как следствие, возникает необходимость в использовании других структур данных и других методах работы с ними. В частности, если для нумерации вершин элементов в узловом МКЭ используются простые методы построения прямой нумерации, то в векторном, как правило, используется глобальная нумерация ребер и соответствующий алгоритм ее построения [4] (см. рисунок).

```
k := 1
Для i = 1, n
    построить список вершин, смежных с i – ой
    для всех e ∈ Ui
        Если ребро (vi, ve) = (ve, vi) ещё не занумеровано,
            то занумеровать его числом k
            k := k + 1
    продолжить перебор по списку Ui
увеличить i
```

Рисунок. Алгоритм построения глобальной нумерации ребер

В современных системах автоматизированного проектирования используется узловой метод конечных элементов с некоторыми допущениями. Дискретизация областей сложной формы производится путем разбиения на треугольники (двумерный случай) или тетраэдры (трехмерный случай).

При создании систем автоматизированных инженерных расчетов используются такие элементы вычислительной математики, как построение разреженных матриц (глобальные матрицы жесткости в SolidWorks / COSMOSWorks и других системах) [7].

Также известно, что в современных системах инженерного расчета используется следующий ряд допущений:

- для каждого конечного элемента при наличии перемещений (углов поворота) в узлах и аппроксимирующей функции рассчитываются деформации. Если элементы линейные, то деформации в пределах элементов постоянные, если элементы параболические, то деформации изменяются линейно;

- при необходимости значения параметров в узлах смежных элементов усредняются с последующим пересчетом напряжений в пределах каждого элемента;
- распределенные нагрузки программно заменяются эквивалентными, сосредоточенными в узлах.

Существуют и другие допущения, более подробно изложенные в специальной литературе, посвященной конкретным программам [1, 7]. Все это приводит к не всегда точным расчетам, а также к часто встречающейся расходимости алгоритмов решения задач.

В памяти компьютера конечно-элементные модели обычно хранятся в векторной форме, т.е. в виде координат совокупности точек, задающих элементы модели. Операции конструирования также выполняются над моделями в векторной форме. В целях уменьшения системных требований в современных САПР выбирается наиболее компактная модель в виде совокупности связанных базовых элементов формы, которая преимущественно и используется для хранения и обработки информации об изделиях в системах конструктивной геометрии [8]. Таким образом, современные САПР используют векторный вид хранения данных для реализации узлового, или скалярного метода конечных элементов.

Перспективы развития алгоритмов на базе МКЭ

Ввиду возрастающих аппаратных мощностей современных серверов и рабочих станций, а также улучшаемого математического аппарата в сфере вычислительной математики полагается возможным разработать комплекс алгоритмов инженерного расчета на основе векторного метода конечных элементов. Очевидно, что в связи с переходом от узлов к ребрам возникает необходимость в построении принципиально новой математической модели, основанной на применении в конечно-элементном анализе теории графов.

Теория графов нашла практическое применение в вычислительной технике в середине 70-х гг. прошлого века. Широкое применение она уже получила при исследовании так называемой проблемы оптимизации, возникающей при конструировании компиляторов. Кроме того, одним из основных направлений в вычислительной технике является построение эффективных алгоритмов и анализ их сложности, и теория графов (а в общем случае, и комбинаторный анализ) предоставляет большие возможности для построения таких алгоритмов [9].

В ходе осуществления большого количества инженерных расчетов в Российском институте радионавигации и времени (ОАО «РИРВ») нами было показано, что даже самые современные системы автоматизированного проектирования и инженерного расчета сегодня не дают необходимой для соответствия стандартам предприятия точности [10]. Вследствие этого возникает необходимость в поиске новых алгоритмов расчета, инновационном прикладном математическом моделировании и оптимизации проектных решений в сфере проведения инженерных расчетов на предприятии.

В связи с вышеизложенными особенностями, достоинствами и недостатками алгоритмов на базе метода конечных элементов, реализованных в современных САПР, в настоящее время создается программа по разработке алгоритмов инженерного расчета на основе векторного метода конечных элементов. Основные этапы реализации программы:

- исследование современных САПР и алгоритмов, в них реализованных;
- исследование последних достижений в сфере вычислительной математики, в частности, векторного МКЭ и возможности его применения в сфере инженерных расчетов радиоэлектронных конструктивов;

- разработка алгоритма конечно-элементной дискретизации простейших конструктивов;
- программная реализация алгоритма конечно-элементной дискретизации.

Данная программа ориентирована на создание методики расчета тепловых режимов радиоэлектронных конструктивов. Методика основана на математических методах теории графов, применяемых в алгоритмах векторного метода конечных элементов.

Литература

1. Алямовский А.А., Собачкин А.А., Одинцов Е.В., Харитонович А.И., Пономарев Н.Б. SolidWorks. Компьютерное моделирование в инженерной практике. – СПб: БХВ-Петербург, 2005.
2. Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений. – М.: Мир, 1984.
3. Писсанецки С. Технология разреженных матриц. – М.: Мир, 1988.
4. Баландин М.Ю., Шурина Э.П. Векторный метод конечных элементов: Учеб. Пособие. – Новосибирск: Изд-во НГТУ, 2001.
5. Рычков С.П. Алгоритмы метода конечных элементов. – СПб: BHV, 2006.
6. Nèdèlec J.C. Mixed Finite Elements in R^3 // Numer. Math.– 1980. – Vol. 35. – №3. – P. 315–341.
7. Алямовский А.А. SolidWorks/COSMOSWorks. Инженерный анализ методом конечных элементов. – М.: ДМК-Пресс, 2004.
8. Норенков И.П. Основы автоматизированного проектирования. – М.: МГТУ им. Н.Э. Баумана, 2002.
9. Свами М., Тхуласимаран К. Графы, сети и алгоритмы: Пер. с англ. – М.: Мир, 1984.
10. Боголюбов Д.А., Кармановский Н.С. Исследование тепловых режимов различных радиоэлектронных конструктивов с помощью системы COSMOSWorks // Научно-технический вестник СПбГУ ИТМО. – 2007. – Выпуск 44. Современные технологии.

АНАЛИЗ АВТОМАТИЗИРОВАННОГО РАБОЧЕГО МЕСТА КАК ОБЪЕКТА ПРОЕКТИРОВАНИЯ

А.С. Федотов

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье производится анализ автоматизированного рабочего места (АРМ). Рассказывается о его назначении, функциях и о составляющих элементах (АРМ). Приводятся примеры программных продуктов, их анализ и качество работы.

Введение

В последние годы возникает концепция распределенных систем управления производством, где предусматривается локальная обработка информации. Для реализации идеи распределенного управления необходимо создание для каждого уровня управления и каждой предметной области автоматизированных рабочих мест (АРМ) на базе профессиональных персональных ЭВМ. Современные масштабы и темпы внедрения средств автоматизации управления в народном хозяйстве с особой остротой ставят задачу проведения комплексных исследований, связанных со всесторонним изучением и обобщением возникающих при этом проблем как практического, так и теоретического характера.

Анализируя сущность АРМ, специалисты определяют их чаще всего как профессионально-ориентированные малые вычислительные системы, расположенные непосредственно на рабочих местах специалистов и предназначенные для автоматизации их работ. Для каждого объекта управления нужно предусмотреть автоматизированные рабочие места, соответствующие их функциональному назначению. Однако принципы создания АРМ должны быть общими: системность, гибкость, устойчивость, эффективность.

Принцип устойчивости заключается в том, что система АРМ должна выполнять основные функции независимо от воздействия на нее внутренних и внешних возможных факторов. Это значит, что неполадки в отдельных ее частях должны быть легко устранимы, а работоспособность системы быстро восстанавливаема. Эффективность АРМ следует рассматривать как интегральный показатель уровня реализации приведенных выше принципов, отнесенного к затратам по созданию и эксплуатации системы. Функционирование АРМ может дать численный эффект только при условии правильного распределения функций и нагрузки между человеком и машинными средствами обработки информации, ядром которых является ЭВМ. Лишь тогда АРМ станет средством повышения не только производительности труда и эффективности управления, но и социальной комфортности специалистов [2].

Структура автоматизированного рабочего места

Автоматизированное рабочее место должно отвечать следующим требованиям:

- своевременное удовлетворение информационной и вычислительной потребности специалиста;
- минимальное время ответа на запросы пользователя;
- адаптация к уровню подготовки пользователя и его профессиональным запросам;
- простота освоения приемов работы на автоматизированное рабочее место и легкость общения, надежность и простота обслуживания;
- терпимость по отношению к пользователю;
- возможность быстрого обучения пользователя;
- возможность работы в составе вычислительной сети.



Рис. 1. Схема автоматизированного рабочего места

Подходы к разработке АРМ

Первый подход – функциональный – представляет собой автоматизацию наиболее типичных функций. Посмотрим, как адаптируется функциональное ПО к конкретным условиям применения. Отметим программные средства, которые являются базовыми при АРМ для различных профессий, связанных с обработкой деловой информации и принятием управленческих решений. Первыми появились программные средства для автоматизации труда технического персонала, что обусловлено, вероятно, большой формализацией выполняемых ими функций. Наиболее типичным примером являются текстовые редакторы (процессоры). Они позволяют быстро вводить информацию, редактировать ее, сами осуществляют поиск ошибок, помогают подготовить текст к распечатке. Применение текстовых редакторов позволят значительно повысить производительность труда машинисток. Специалистам часто приходится работать с большими объемами данных, чтобы найти требуемые сведения для подготовки различных документов. Для облегчения такого рода работ были созданы системы управления базами данных (СУБД: DBASE, RBASE, ORACLE и др.). СУБД позволяют хранить большие объемы информации и, что самое главное, быстро находить нужные данные. Так, например, при работе с картотекой постоянно нужно перерывать большие архивы данных для поиска нужной информации, особенно если карточки отсортированы не по нужному признаку. СУБД справится с этой задачей за считанные секунды. Хотя современное ПО отвечает почти всем требованиям, налагаемых на него работниками различных профессий, чего-то все равно всегда не хватает. Поэтому большим плюсом такого ПО является возможность его доработки и изменения. Что же касается разработки новых программных средств в АРМ, то она ведется по двум направлениям: создание нового ПО для новых профессий и специализация ПО для уже существующих профессий. В настоящее время наблюдается тенденция перехода к созданию АРМ профессионального назначения. Оно выражается в следующем:

- учет решаемых задач;
- взаимодействие с другими сотрудниками;
- учет профессиональных привычек и склонностей.

Оснащение такими АРМ позволяет повысить производительность труда учреждений работников, сократить их численность и при этом повесить скорость обработки экономической информации и ее достоверность, что необходимо для эффективного планирования и управления [1].

Состав программно-технических средств вычислительной техники комплексов АРМ

Методы организации программно-технических средств в комплексах АРМ должны определяться в общем контексте рассматриваемых процессов оперативного управ-

ления производством (ОУП) промышленных предприятий, целевая функция которых – минимизировать затраты всех видов ресурсов на изготовление установленной номенклатуры. Синтез методов и моделей организации программно-технических средств при представлении АС оперативного управления производством как комплексов АРМ хозяйственных производственных коллективов должен пройти две стадии: стадию определения рационального состава средств ВТ и стадии решения проблемы распределения ресурсов вычислительной системы комплексов АРМ по конечным ее потребителям [5].

При определении рационального состава средств ВТ и базового программного обеспечения значимо затрагиваются три вида ресурсов – трудовые, временные и материальные. Причем затрагиваются как ресурсы заказчика, так и ресурсы разработчика. Все переменные представляют собой совокупность комбинированных сочетаний, доступных заказчику для приобретения средств ВТ и базового программного обеспечения. Распределим эти переменные по двум группам – характеристики комбинаций, существенные для обеих сторон развития процесса, т.е. и для разработчика, и для заказчика, и существенные только для разработчика.

Техническая (аппаратная) совместимость новых средств ВТ по отношению к существующему парку ВТ у заказчика и к парку ВТ, прогнозируемому для приобретения в дальнейшем. Практика показывает, что этот показатель – один из важнейших принимаемых во внимание при выборе ВТ. Тенденция приобретения средств ВТ, аппаратно совместимых с имеющимися, связана со многими объективными и субъективными причинами, где не последнее место занимает и психология заказчика, его чувство уверенности в успехе применения именно этого класса аппаратных средств.

Программная совместимость, которая определяется совместимостью аппаратно-реализованной системы команд, совместимостью форматов представления данных, совместимостью трансляторов и т.д. Значительное влияние этого показателя на расход ресурсов может объясняться наличием больших объемов ранее подготовленных нормативных, архивных и статистических данных, а также специализацией подготовленного персонала на предприятии, имеющего опыт работы с конкретными базовыми средствами программного обеспечения.

Эксплуатационная совместимость внутри приобретенного комплекса средств ВТ, которая позволяет в случае выхода из строя отдельных модулей АРМ либо оперативно производить замену вышедшего из строя модуля, либо провести переназначение используемых устройств между конкретными АРМ в пределах вычислительных ресурсов всех комплексов (внутри цехового комплекса, внутри межцехового комплекса, внутри системы всякого предприятия).

Надежность средств ВТ по техническим условиям и ее соответствие конкретным условиям работы – вибрации, окислению, пыли, загазованности, скачкам напряжения и т.д., что требует дополнительных средств защиты.

Совокупная скорость решения функциональных задач по видам АРМ-комплекса – скорость обработки существующих объемов данных в различных режимах работы. Обычно для определения значений этого показателя недостаточно знать только объемы информационной базы конкретного АРМ и паспортные характеристики и предоставляемых вычислительных ресурсов.

Наиболее существенные показатели, влияющие на затраты трудовых и временных ресурсов и разработчика и заказчика при использовании конкретной конфигурации программно-технических средств комплексов АРМ и влияющие на их выбор:

- стоимость реализации «дружественного интерфейса» включает и программы обучения, и возможность получения справок в процессе работы на АРМ о способах продолжения или окончания диалога;
- возможность изменения состава и содержания реализуемых на конкретных АРМ функций, в том числе перераспределения между персоналом;

- обеспечение требований защиты от несанкционированного доступа для баз знаний и баз данных, а также их «прозрачности» в случае необходимости;
- стоимость обеспечения программно-аппаратных связей во внутрицеховых комплексах АРМ, между внутрицеховыми комплексами АРМ с другими подразделениями предприятия;
- обеспечение указанных видов связей, кроме этого, может быть также детализировано в соответствии с терминологией, принятой для локальных вычислительных сетей, а именно: передача данных, эмуляция терминалов, доступ к удаленным ресурсам, включая доступ к данным и запуск задач. Особо может быть выделен показатель поддержки стандартных протоколов обмена в сетях ЭВМ;
- возможность расширения комплексов ВТ за счет наращивания технических устройств (блоками оперативной памяти, внешними запоминающими устройствами, дополнительными процессорами, принтерами и т.д.).

Вторая группа показателей, существенно влияющих при выборе средств ВТ и базового программного обеспечения на затраты временных и трудовых ресурсов только разработчика, состоит из небольшого числа характеристик. Внешне эта группа выглядит как определяемая для удовлетворения показателей первой группы, как их своеобразная поддержка, хотя и имеет свои особенности [3]:

- скорость создания программного продукта в качестве обеспечения баз знаний и комплексов АРМ. Оценка показателя может быть получена экспертным путем, на основе опыта по изготовлению аналогов, с учетом предполагаемой технологии проектирования систем и уровня квалификации коллектива исполнителей;
- ожидаемые затраты на создание программного продукта в качестве обеспечения баз знаний комплексов АРМ. Стоимостная оценка затрат разработчика в части расхода всех видов ресурсов может определяться как по предыдущему показателю, так и экспертным путем;
- удобство сопровождения созданного программного продукта. Показатель в данном случае – статистический, определяемый группой факторов, таких, как возможность изменение отдельных элементов АРМ, не затрагивая общего математического обеспечения комплекса АРМ, скорость и затраты на подобного рода работы;
- полнота обеспечения стандартными средствами защиты данных и материальное обеспечение требований конечного потребителя и разработчика.

Выбор оптимального решения возможен для различных вариантов учета показателей. В полном случае учитываются все показатели, в частном случае – часть только наиболее важных. Представление свободы выбора позволяет определять выбор конкретной комбинации средств ВТ и базового программного обеспечения в условиях частичного отсутствия знаний о них, а кроме этого, позволяет снижать размерность задачи, используемой для решения модели в целях повышения эффективности использования средств ВТ.

Экономико-организационные резервы производства программных средств АРМ

Современные масштабы работ по производству программных средств (ПС) вычислительной техники как продукции производственно-технического назначения требуют незамедлительного решения одного из важных вопросов в данной области – резкого повышения производительности труда персонала, занятого в разработке ПС. Это обусловлено тем, что за последние десятилетия производительность труда работников, связанных с программированием, возросла примерно в 2–3 раза, тогда как производительность ВТ повысилась ориентировочно на два порядка. С другой стороны, в ближайшее время можно ожидать скачка в росте спроса на заимствуемые ПС, особенно

пакеты прикладных программ, связанного с широким внедрением в народном хозяйстве автоматизированных рабочих мест, в том числе на основе персональных ЭВМ.

Анализ предлагаемых решений по устранению складывающейся диспропорции показывает, что их можно объединить в два направления:

- экстенсивное, связанное с расширением сети организаций по разработке, производству и поставке ПС;
- интенсивное, но затрагивающее лишь вопросы роста производительности труда разработчиков ПС за счет тех источников, которые по аналогии с другими отраслями можно отнести к техническим факторам, т.е. применение современной технологии программирования и автоматизированных инструментальных средств разработки программ.

Первым обращает на себя внимание то, что в качестве ближайшего отраслевого аналога для создания ПС избирается машиностроение. Основанием для этого принято считать то, что ПС рассматривается как специфицированное изделие. Опасность заключается в желании саму организацию работ по созданию ПС провести по образцу и подобию прототипов из машиностроения. Причем предпочтение отдается крупносерийному и серийному производству, тогда как преобладающая серийность тиражирования ПС в десятках, редко сотнях или тысячах экземпляров скорее подсказывает необходимость подойти к ним как к изделиям мелкосерийного, если не единичного производства. Но главным, на что следует обратить внимание в первую очередь, являются различия в характере труда работников машиностроения и разработчиков ПС, которые существуют в настоящий момент и, скорее всего, сохранятся в будущем. Преобладающее в машиностроении пооперационное разделение труда по стадиям техпроцесса с организацией подразделений по этим стадиям в создании ПС не проходит проверку широкой практикой, так как умственный труд по программированию гораздо сложнее разбить на последовательность формализуемых операций, нормирование которых достаточно достоверно учитывает и характеристику персонала, и специфику объекта внедрения норм.

Если ставить вопрос об аналогах, то гораздо ближе к разработке ПС по характеру труда – научно-исследовательские работы и работы опытно-конструкторского характера (НИОКР). С учетом сложности выделения, формализации и нормирования отдельных операций по созданию ПС наиболее целесообразной с точки зрения установления хозрасчетных отношений и стимулирования труда разработчиков ПС следует признать организацию их работы на основе заказ-нарядов с расчетами за полностью сданную программную продукцию. Основной организационной единицей в таком случае может являться бригада, специализированная на выполнении цикла работ по выпуску либо готового изделия в объеме пакета прикладных программ среднего размера, либо отдельного компонента сложного программного комплекса.

Судя по литературе, существует достаточно большой положительный опыт бригадной разработки ПС, однако, к сожалению, экономические аспекты деятельности таких бригад не рассматриваются. Учитывая объективную слабость пооперационной нормативной базы программирования, следует признать весьма привлекательной аккордную форму оплаты труда бригад, тем более что в настоящее время уже складывается достаточно достоверная нормативная база по укрупненным работам создания ПС [4]. Применение аккордной системы, безусловно, стимулирует сокращение затрат и сроков разработки ПС в гораздо большей степени, нежели преобладающая в настоящее время в этой области система должностных окладов. Оценку результатов работы обеспечивающих подразделений можно в таком случае вести по объему предоставляемых услуг, а не процентом отчислений от суммы заказ-нарядов, что явно не стимулирует обеспечивающие подразделения на рост объемов и качества услуг. Планирование объемов и номенклатуры услуг должно вестись на основе внутренних планов-соглашений

(договоров) между подразделениями основного и вспомогательного производства ПС. Таким же образом можно провести оценку результатов использования одним подразделением-разработчиком ПС программной продукции другого подразделения (передача программных модулей). В основу оценки взаимосвязи подразделений следует, по-видимому, положить нормативные (не пооперационные) затраты времени на разработку программ и услуги. От этих затрат возможен в достаточной степени достоверный переход к стоимостным оценкам через заработную плату и систему коэффициентов в ней, учитывающую прочие расходы. Введение таких отношений во всех организационных уровнях разработки ПС, включая и низовые, позволит мобилизовать противозатратный механизм в их производстве, что чрезвычайно важно при существующем порядке ценообразования на ПСВТ, когда цена (Ц) рассчитывается по формуле, близкой к традиционной:

$$Ц = С + Пн + Пд ,$$

где С – себестоимость разработки (разовых) и производства (тиражируемых) ПС; Пн – нормативная прибыль, устанавливаемая централизованно, в процентах к С; Пд – дополнительная прибыль, устанавливаемая в зависимости от величины экономического эффекта при эксплуатации ПС.

Таким образом, нельзя не заметить прямого стимулирования затрат в ценообразовании ПС, когда и разработчик, и производитель заинтересованы в росте величины С, чего тем более не особенно сложно достигнуть при отсутствии общих нормативов в разработке ПС. Внедрение внутрипроизводственных отношений на уровне бригад (отделов, цехов) может играть роль в снижении величины С. Тогда предоставляемые им средства на разработку программ фиксированы, а требуемые затраты ресурсов (особенно трудовых и машинных) определяются этими единицами с предоставлением им большей части полученной экономии в стоимостном выражении, что близко к условиям аккордно-премиальной оплаты труда. Попутно заметим, что только тогда становится возможным формирование реальной нормативной базы на разработку ПС, так как нормативы должны формироваться путем анализа опыта труда персонала, заинтересованного в интенсивной работе, что в настоящее время существует в очень малой степени.

Программный комплекс LCAD

Методом разработки автоматизированного рабочего места проектировщика технологических систем можно считать метод использования различных программных комплексов. Программный комплекс LCAD (от Layout CAD – расстановка оборудования с помощью компьютера) предназначен для создания автоматизированного рабочего места проектировщика, осуществляющего технологическое проектирование новых производственных помещений (рис. 2), а также технологическую реорганизацию существующего производства. Комплекс может быть также использован для получения различной справочной информации по установленному на производстве и введенному в базу данных системы оборудованию.

Программный комплекс LCAD позволяет автоматизировать процесс формирования:

- строительной подосновы (планов этажей зданий) по одноэтажным и многоэтажным промышленным помещениям для последующего размещения технологического оборудования; а также административно-бытовым зданиям;
- графической и текстовой документации по технологической планировке производственных помещений.

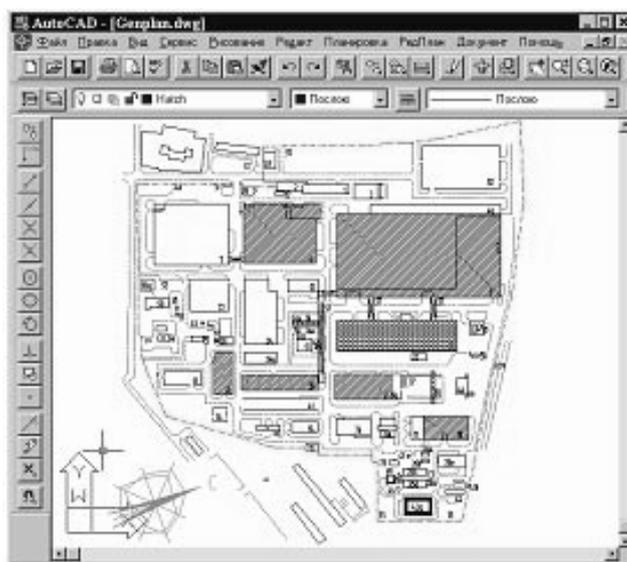


Рис. 2. Проектирование новых производственных помещений

LCAD обеспечивает создание и ведение базы данных (БД), содержащей массивы текстовой и графической информации. Структура массивов БД позволяет загружать и использовать при проектировании следующие виды информации:

- характеристики оборудования (наименование и модель, габариты, масса, установленная мощность электродвигателя и некоторая дополнительная информация), с обеспечением поиска и выбора информации по классам и группам оборудования;
- дополнительная графическая информация по оборудованию – размеры, установочные планы, планы опор, точки подключения электропитания, воздуха и т.п.;
- темплеты ("габаритки", "фишки") оборудования;
- спецификации по установленному оборудованию;
- принятые условные графические обозначения для нанесения на планировки;
- структура производства (промышленная площадка – производственный корпус – цех – участок);
- генплан предприятия (для обеспечения быстрого выхода на нужную планировку производственных корпусов, цехов, участков);
- любая информация по цехам и участкам предприятия (виды и размеры площадей и т.д.);
- справочные данные по нормам и требованиям к размещению оборудования.

LCAD предполагает создание и хранение в БД технологических планировок на строительной подоснове производственного корпуса (здания) в целом. Спецификация установленного оборудования (рис. 3) создается и хранится в БД в целом по предприятию. Оформление и вывод на печать графической (чертежи планировок) и текстовой (спецификации оборудования) документации может производиться как в целом по производственным корпусам, так и по отдельным цехам и участкам, запрашиваемым в БД.

LCAD использует и расширяет возможности пакета AutoCAD фирмы Autodesk за счет наличия дополнительного набора специальных приложений, обеспечивающих основные функции проектирования технологических планировок цехов и участков предприятия [3].

Комплекс можно использовать в технологических подразделениях и технических отделах как крупных предприятий, так и небольших производственных организаций, применяющих автоматизированные рабочие места технологов-проектировщиков на базе персональных компьютеров.

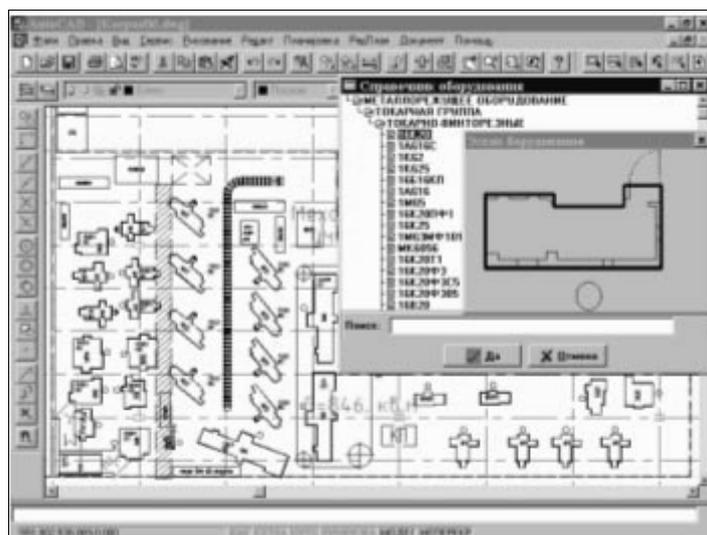


Рис. 3. Спецификация установленного оборудования

Заключение

Анализируя сущность АРМ, специалисты определяют их чаще всего как профессионально-ориентированные малые вычислительные системы, расположенные непосредственно на рабочих местах специалистов и предназначенные для автоматизации их работ. Для каждого объекта управления нужно предусмотреть автоматизированные рабочие места, соответствующие их функциональному назначению. Однако принципы создания АРМ должны быть общими: системность, гибкость, устойчивость, эффективность.

Литература

1. Силкин В. Интеллектуализация электронных устройств // Компоненты и технологии – 2005. – №3. – С.37–39.
2. Автоматизированное рабочее место в системе управления предприятием. Сборник научных трудов. – Л., 1989.
3. Шураков В.В. Автоматизированное рабочее место для статической обработки данных. – 1990.
4. Кантарь И.Л. Автоматизированные рабочие места управленческого аппарата. – 1990.
5. Комплексная автоматизация проектно-конструкторских и технологических работ. [Электронный ресурс]. – Режим доступа: www.cad.ru

ВОЛНОВОЙ АЛГОРИТМ ВЕКТОРИЗАЦИИ ЛИНЕЙНЫХ РАСТРОВЫХ ИЗОБРАЖЕНИЙ

С.В. Москаленко

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье рассмотрено новое решение проблемы векторизации бинарных растровых изображений, обладающее повышенной эффективностью. Под эффективностью понимается повышенное быстродействие и качество выходного изображения.

Введение

Достигнутые результаты в области цифровых методов представления и сжатия изображений на сегодняшний день не решают в полной мере задачу их эффективного представления, что делает поиск такого представления актуальным. Одним из путей решения представляется разработка интеллектуальных технологий обработки изображений, обеспечивающих не формализованное их кодирование, а распознавание их пространственной структуры, которая и является носителем заключенной в них информации, поскольку возможности извлечения полезной информации из изображений целиком и полностью определяется их пространственно-структурными свойствами.

На данный момент существует множество алгоритмов автоматической трассировки (векторизации) растра. Анализ выявил недостаточную надежность данных алгоритмов с точки зрения выходных характеристик (быстродействие, качество выходного изображения).

В статье рассмотрен подход, который обладает лучшими выходными параметрами за счет оптимизации основного алгоритма. При векторизации происходит отсеивание лишней информации, сохраняются информативные данные для последующего этапа распознавания образов.

Анализ существующих подходов

Задача распознавания может состоять из нескольких шагов. Одним из начальных будет являться этап перехода от растрового к векторному изображению. Под векторным представлением растрового изображения будем понимать такой нагруженный граф, у которого нагрузка вершин – пары координат x , y соответствующих узловых точек изображения. На данный момент существует множество алгоритмов трассировки (векторизации) растра. Их условно можно разделить на две группы: автоматической векторизации, когда происходит трассировка всего растрового изображения без вмешательства пользователя (все необходимые параметры задаются до запуска), и автоматизированной векторизации, когда весь процесс трассировки сопровождается корректировками пользователя. К первой группе относятся такие программные продукты, как CorelTrace Centerline, CorelTrace Advanced Outline (входят в состав пакета Corel Draw 12), AutoTrace Outline программы AutoTrace.

В данной статье упор сделан на автоматические алгоритмы векторизации. Был проведен их анализ по основным характеристикам, таким как скорость работы, процент потери важных информативных данных входного растра, процент искажения от первоначального изображения, целостность распознанных элементов, эффективность кодирования примитивов, наличие функции восстановления искажений и соответствие типов линий (входных–выходных).

Алгоритм CorelTrace Centerline обладает высокой скоростью работы, небольшим процентом искажения от первоначального растра. Данный алгоритм разбивает растр на примитивы по принципу: примитив (вектор или полилиния) считаются сформирован-

ными, если на пути их трассировки встретилось ветвление. После обнаружения ветвления трассировка идет по найденным ветвям, образуя новые примитивы. Данный алгоритм работает по принципу вычисления средней линии путем предварительного истончения раstra (рис. 1, б). И, значит, к недостаткам CorelTrace Centerline можно отнести, во-первых, потерю информации о контуре растрового примитива (остаются только координаты средней линии), а соответственно, и о типе вектора. Во-вторых, необходим ввод дополнительного этапа, реализующего истончение растрового изображения.

Еще один алгоритм, CorelTrace Advanced-Outline 12, в отличие от предыдущего метода, только оконтуривает растр (рис. 1, в). Это само по себе не является эффективным представлением примитива, так как в идеале выходной вектор должен определяться лишь координатами начала, конца и своим типом.

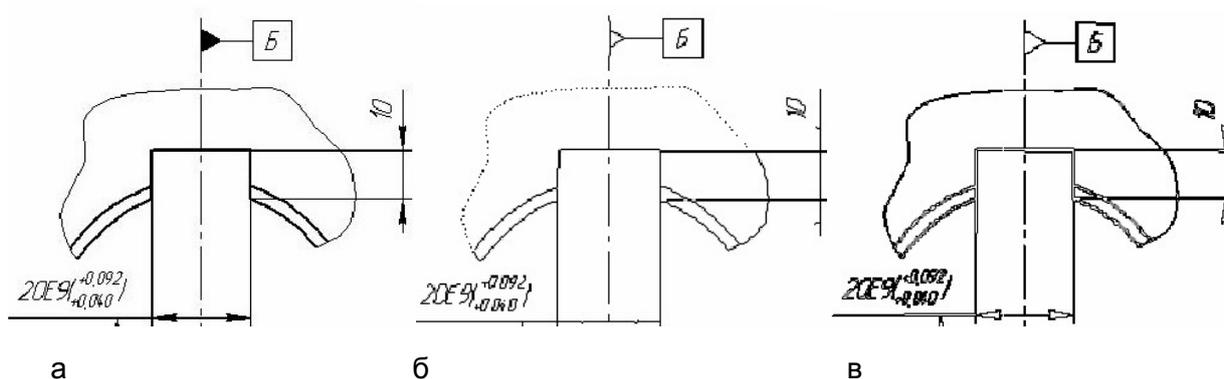


Рис. 1. Примеры работы векторизаторов: а – исходное изображение; б – CorelTrace Centerline; в – CorelTrace Advanced-Outline

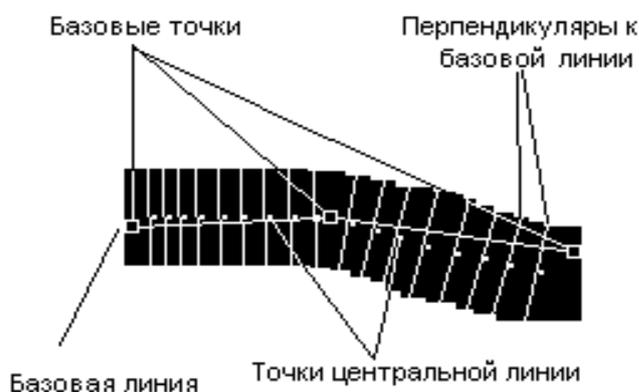


Рис. 2. Вычисление точек средней линии растрового изображения

Существуют подходы, объединяющие достоинства двух описанных ранее алгоритмов. Происходит расчет средней линии одновременно с анализом контура [1]. При получении координат точки запускается процесс, состоящий из следующих этапов: центрирование точки указания, определение направления (путем построения циклической гистограммы), вычисление точек центральной линии. После определения предпочтительных направлений производится замер толщин (ширина линии в пикселях) в направлениях, перпендикулярных двум базовым, одновременно с этим происходит процесс вычисления точек центральной линии в обоих направлениях с шагом в одну растровую точку. Процесс замера толщин контролирует резкое изменение толщины, что

означает наличие утолщения, утоньшения или пересечения. При обнаружении вышеописанной ситуации процесс останавливается в данном направлении и начинается в противоположном (см. рис. 2). Попиксельный расчет толщин является весьма трудоемкой операцией. Также для работы алгоритма необходимо содействие пользователя в выборе определенных констант, таких как максимальная ширина линии и допустимый разрыв линии, что лишает систему автоматического режима работы.

Реализация алгоритма

В разработанном алгоритме векторизации отсутствует как попиксельное вычисление толщины, так и этап определения предпочтительных направлений. Сразу после ввода изображения запускается побитовое сканирование этого изображения на наличие «черных» точек раstra. После получения исходной точки от нее в основных направлениях строится циклическая гистограмма по черным пикселям до первого белого. Гистограмма строится с определенным угловым шагом, шаг был взят равным пяти градусам. Для вычисления координат точек при построении гистограммы используется типовой алгоритм Брезенхема (генерации отрезка). В результате могут получиться гистограммы следующих типов: с ярко выраженными пиками либо же неопределенные.

Под ярко выраженным пиком понимается участок на графике, функция которого резко возрастает, а потом убывает. При этом участок должен быть достаточно узок. Так, например, если график будет иметь один или два пика, это будет говорить о том, что исходная точка принадлежит отрезку или дуге, т.е. линейному примитиву (рис. 3, а). Если на функции обнаружатся более двух пиков, это будет значить, что исходная точка лежит на пересечении линейных объектов (рис. 3, в). Неопределенной будем считать ту гистограмму, которая имеет недостаточно узкие пики (рис. 3, г). Такие пики суть отображение площадных объектов на растре. Проанализировав функции, можно отнести существующие объекты растрового изображения либо к площадным, либо к линейным.

Следующий шаг после построения гистограммы – определение толщины вектора (для линейных объектов). Если найдено только два явных пика, то можно сделать вывод, что исходная точка достаточно удалена от обеих вершин линейного объекта. Если же пик один, то точка примерно совпадает с одной из вершин. Для случая нахождения двух явных пиков должно быть вычислено среднее значение величин на гистограмме, лежащих в непосредственной локальной близости от точки экстремума, принадлежащей участку функции между двумя вершинами пиков (только одна точка экстремума может быть найдена). Полученное значение будет отражать толщину линии (рис. 3, а).

Для случая одного явного пика должно быть вычислено среднее значение величин, лежащих в окрестности точки перехода от выпуклого участка функции к вогнутому (или прямолинейному), которая может находиться как слева, так и справа от вершины пика. Полученное значение также будет отражать толщину линии (рис. 3, б).

Проанализировав гистограмму, мы будем знать толщину линии, координаты двух концов поперечного сечения и координаты исходной точки. Теперь можно найти точку центра масс на данном сечении – это будет его середина. Чтобы организовать отслеживание центральной линии и «линейность» растрового объекта, реализован волновой алгоритм. Суть его заключается в построении сферической волны, центр которой совпадает с последним найденным центром масс и радиусом, равным ширине растрового объекта в исходной точке.

Отцентрировав исходную точку на сечении, алгоритм генерирует окружность радиусом равным толщине линии, используя типовой алгоритм Брезенхема (генерации окружности). Построенная окружность должна пересечь границу растрового объекта с фоном $2n$ раз, в зависимости от результатов анализа построенной ранее гистограммы.

Например, если сделан вывод, что исходное сечение не совпадает с вершиной линейного объекта, значит, окружность должна пересечь границу четыре раза, так как растровый объект будет развиваться в двух направлениях.

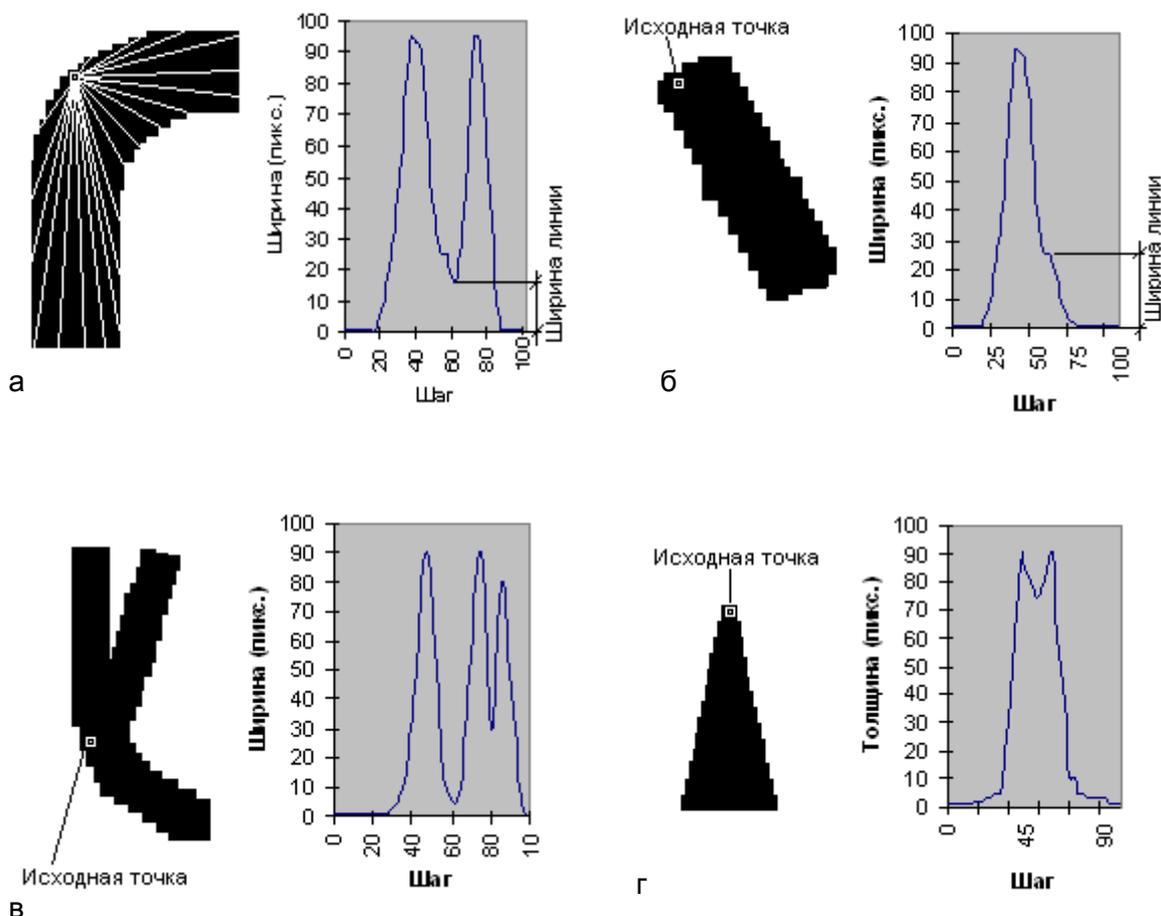


Рис. 3. Примеры гистограмм растровых изображений (слева – растровое изображение, справа – гистограмма): а – дуга; б – исходная точка совпадает с вершиной прямой; в – исходная точка лежит на пересечении; г – площадной объект

Далее проведем через пары точек две прямые таким образом, чтобы образовалось два продольных сечения. Каждая из середин полученных сечений будет являться новым центром масс, из которого будет выходить очередная волна. Итерационное развитие волн в одном направлении будет продолжаться до тех пор, пока окружность не пересечет растровый объект в количестве точек, не равном двум, либо же не произойдет резкое изменение ширины продольного сечения, что будет говорить о наличии утоньшения, утолщения или пересечения (рис. 4). При обнаружении вышеописанной ситуации процесс останавливается в данном направлении и начинается в противоположном. Таким образом, данный алгоритм получит координаты точек центральной линии с информацией о размере сечения в каждой. Эти параметры необходимы для векторного представления линейных растровых объектов.

Что касается генерации самой волны, то во время первой итерации необходимо строить полную окружность, а для последующих итераций можно ограничиться лишь разверткой $(9/8)\pi$. Развертка должна откладываться по обе стороны от перпендикуляра, построенного к сечению, которое было получено в результате предыдущего построения волны.

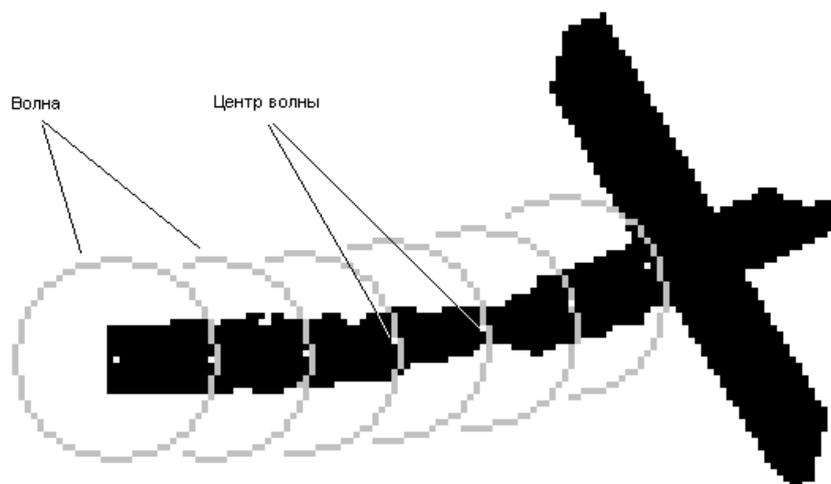


Рис. 4. Вычисление точек средней линии растрового объекта

Заключение

Разработанный алгоритм, предназначенный для обработки и отображения графических данных, обладает повышенной эффективностью, причем под эффективностью понимается не только повышение быстродействия алгоритма, но и улучшение качества получаемого решения. Эта эффективность обусловлена использованием генерации сферических волн для нахождения центральных линий растровых объектов. К тому же анализ гистограмм позволяет добиться автоматической работы алгоритма без вмешательства человека.

Преобразование изображения к векторному виду дает возможность редактирования и манипулирования распознанных объектов.

Литература

1. Крылов А.Б. Способ выбора растровых объектов на монохромном изображении с автоматическим вычислением геометрических параметров // Моск. гос. техн. ун-т им. Н.Э. Баумана – 2001. – Интеллектуальные технологии и системы. Вып. 5.
2. Анисимов Б.В., Курганов В.Д. Злобин В.К. Распознавание и цифровая обработка изображений: Учеб. пособие. – М.: Высш. шк., 1983. – 295 с. ил.
3. Горелик А.Л., Скрипкин В.А. Методы распознавания: Учеб. пособие. – 2-е изд., перераб. и доп. – М.: Высш.шк., 1984. – 208 с., ил.
4. Дуда Р., Харт П. Распознавание образов и анализ сцен. – М.: Мир, 1976.
5. Прэтт У. Цифровая обработка изображений (в 2-х книгах). – М.: Мир, 1982.
6. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978.
7. Фу К. Последовательные методы в распознавании образов. – М.: Наука, 1971.
8. Фу К. Структурные методы в распознавании образов. – М.: Мир, 1977.
1. Хант Э. Искусственный интеллект. – М.: Мир, 1978.
9. Барабаш Ю.Л. Учет свойств признаков при распознавании // Известия АН СССР. Техническая кибернетика. – 1965. – №5.
10. Чечкин А.В. Математическая информатика. – М.: Наука. Гл. ред. физ.- мат. лит., 1991.
11. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: Изд-во института математики, 1999.

12. Горелик А.Л., Гуревич И.Б., Скрипник В.А. Современное состояние проблемы распознавания. – М.: Радио и связь, 1985. – 160 с.
13. Журавлев Ю.И. Об алгебраическом подходе к решению задач распознавания и классификации // Проблемы кибернетики. – М.: Наука, 1978. – Вып. 33. – С. 5–68.
14. Гренандер У. Лекции по теории распознавания образов. – М.: Мир, 1979, 1 том; 1981, 2 том; 1983, 3 том.
15. Жданов А.А. Об одном имитационном подходе к адаптивному управлению // Вопросы кибернетики. – Выпуск 2. – М., 1996. – С. 171–206.
16. Жданов А.А. Метод автономного адаптивного управления // Известия АН. Теория и системы управления. – 1999. – № 5. – С. 127–134.
17. Образовательный портал «Распознавание образов и искусственный интеллект» [Электронный ресурс]. – Режим доступа: <http://www.ocrai.narod.ru>, свободный. – Загл. с экрана. – Яз. рус.

ВОЗМОЖНОСТИ ГРАФ-ОРИЕНТИРОВАННОГО ПРОГРАММИРОВАНИЯ ДЛЯ СИСТЕМ АВТОМАТИЗИРОВАННОГО УПРАВЛЕНИЯ ПРОЦЕССАМИ НА ПРИМЕРЕ jPDL ЯЗЫКА

В.В. Власов

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье рассматриваются базовые принципы и модели граф-ориентированного программирования на примере языка jPDL (java process definition language), описаны основные отличия от объектно-ориентированного подхода, приведен пример классов, диаграмм и описания процесса, а также объясняется логика выполнения процесса на основе движения «маркера выполнения» по траектории графической диаграммы

Введение

Существующие на сегодняшний день объектно-ориентированные языки программирования не позволяют разрабатывать гибкие, легко (или достаточно легко) настраиваемые, а также легко изменяемые приложения (имеется в виду бизнес-логика). А о том, чтобы не профессионал мог изменять ход выполнения программы, еще несколько лет назад не могло быть речи. Это особенно важно в таких областях, как например, управление бизнес-процессами предприятия, в системах автоматизированного управления ресурсами предприятия, рабочими процессами, документооборота. Даже такой огромный шаг в этом направлении, как появление кросс-платформенного языка Java от Sun Microsystems, не решает эту проблему. Это связано не с тем, что языки программирования плохо продуманы и реализованы, а с тем, что модель объектно-ориентированного языка программирования не включает в себя некоторые важные аспекты, которые интуитивно понятны человеку – поддержка сохраняемого состояния ожидания и графическое представление. Поэтому в последнее время появляется большое количество готовых решений, каркасов (frameworks) для упрощения процесса разработки, позволяющих уменьшить время, затрачиваемое разработчиком на изменение приложения в соответствии с постоянно изменяющимися бизнес-требованиями. Таким образом, уменьшается время между выходами новых релизов продуктов.

Граф-ориентированное программирование (GOP, graf oriented programming) сейчас только оформляется как направление в программировании, хотя общие принципы известны достаточно давно. Типичный представитель языка GOP – Java язык описания процессов jPDL[1], который и будет рассмотрен в данной статье. Разработчики этого языка создали для него и движок выполнения процессов (process virtual machine), который позволяет гибко использовать любой процессный язык. Этот язык интересен еще и тем, что он реализует основные принципы граф-ориентированного программирования без привязки к какой-то конкретной реализации. Поэтому он является хорошим примером как теоретического, так и прикладного применения принципов GOP.

Для удобства в данной статье будут сохранены все английские названия на случай, если читатель захочет самостоятельно продолжить разбираться в вопросах граф-ориентированного программирования и jPDL.

Языки предметных областей

Любой язык описания процессов может быть отнесен к классу Domain Specific Language (DSL)[2], языков предметных областей. Важным аспектом DSL является наличие определенной грамматики языка. Так, для языка программирования Java базовыми являются понятия класса, метода, поля, конструктора, для правил – условие, резуль-

тат, для GOP – это узел графа, переходы, события (действия). Основная идея DSL заключается в том, что разработчик «думает» в терминах данной грамматики, когда создает компоненты (или программы), используя этот язык. Так, среды разработки (IDE) для каждого языка программирования строятся вокруг определенных терминов (грамматик) этого языка. Конечно же, кроме вербального описания языка, существует его конкретная реализация в виде файла кода программы (текстовый), блок-схемы (графический). Хотя в объектно-ориентированном программировании и существует практика создания блок-схем на этапе проектирования (или других графических представлений логики, например, с использованием нотаций UML), но, тем не менее, в дальнейшем при кодировании (компилировании, выполнении), т.е. непосредственно при написании кода, создания выполняемых компонентов или во время выполнения программы блок-схемы не являются частью выполнения этих процессов. Для Java разработчик пишет текстовый код программы (файл java) – «исходник», который потом компилируется в byte-код, в виде class-файла, который и является выполняемым. Для DSL характерно существование нескольких представлений «кода», или инструкций для выполнения.

Например, для jPDL – это XML-файл описания процесса (который содержит наборы команд и помещается в исполняемую среду) и его графическое представление, которое, в отличие от блок-схем, является важной и очень значимой частью на этапе «кодирования».

Соответственно для GOP, представителем которого является jPDL, логично существование двух редакторов – текстового и графического. Пока разработчик рисует графовую модель процесса, генерируется XML-представление этого процесса с использованием соответствующих нотаций. И это не является роскошью – это просто новый шаг вперед в гибкой разработке ПО, когда значительная часть рутинной работы перекладывается на интегрированную среду разработки (IDE). Однако в любой момент разработчик может отредактировать XML-представление. Современные среды разработок (например, свободно распространяемый Eclipse) осуществляют двунаправленную связь между графическим и текстовым представлениями.

Если десять лет назад значительная часть времени работы программиста тратилась на кодирование, то сейчас происходит смена приоритетов в направлении именно DSL языков. Эта смена приоритетов связана также с появлением большого числа готовых решений, каркасов (frameworks) – jPDL для потоков работ (workflow) в java [1], BPEL для оркестрации веб-сервисов (web-service orchestration) [3], SEAM для pageflow [4]. GOP – это общий фундамент для всех этих типов DSL (рис.1).

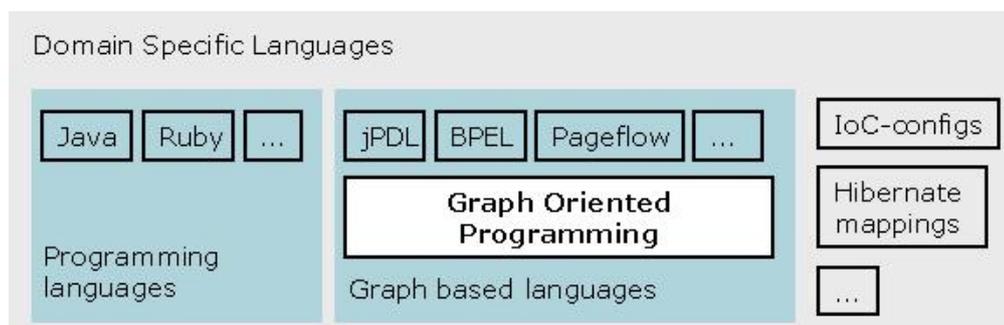


Рис. 1. Положение граф-ориентированных языков

В объектно-ориентированном программировании (ООП) структурирование программы происходит группированием методов с данными. Аспект-ориентированное программирование (АОП) [5] предлагает путь извлечения взаимопересекающихся соотношений, а такие модели, как внедрение зависимости (Dependency Injection – DI) [6]

и инверсия контроля (Inversion of Control – IoC)[6], могут облегчить разводку (трассировку) объектов графа, однако в GOP используются другие подходы, о которых будет говориться в статье. Более подробные сведения о DSL языках можно получить в [7].

Возможности GOP

Существует достаточно большое количество языков описания процессов, основанных на графовой модели. Однако между ними есть большая разница, в том числе и в области применения: так, BPEL – это язык, основанный на XML, и используется для «оркестрации» сервисов в архитектуре с общей шиной, Enterprise Service Bus (ESB) [3]. Pageflow process language, язык управления навигацией страниц, используется для веб-приложений. Это два примера использования процессных языков в различных областях.

Несмотря на эти различия, существуют, как минимум, два аспекта, которые обнаруживаются в любом процессном языке: поддержка состояния ожидания и графическое представление. И это не просто совпадение. Дело в том, что это как раз те возможности, которые недостаточно хорошо поддерживаются в простых ООП, таком как Java.

GOP как бы дополняет ООП этими двумя возможностями. Зависимость GOP от ООП означает, что все конкретные процессные языки, реализуемые на основе GOP, должны будут разрабатываться в среде определенного ОО языка. Правда, это относится далеко не ко всем языкам. Язык jPDL, например, предоставляет такую возможность за счет Process Virtual Machine (виртуальная машина процессов) [1], в отличие от языка BPEL, который является только языком описания процессов и никак не привязывается к какому-либо языку программирования.

Состояние ожидания

Императивные языки программирования (например, как Java) используются для определения последовательности инструкций, которые должны быть выполнены системой. Это прекрасный язык для описания таких процессов, как механизм «запрос-ответ» на сервере. Система непрерывно выполняет последовательности инструкций, записанные в коде, начиная от получения запроса и до отправки ответа.

Однако такой механизм обычно является частью более сложного сценария. Например, клиент отправляет заказ на покупку. Этот заказ должен быть проверен и подтвержден менеджером, после чего информация вносится в ERP-систему. Этот процесс сопровождается созданием задач для персонала и помещением их в списки задач для каждого сотрудника, причем для разных процессов могут быть определены разные ответственные лица. В конце концов подтверждение отправляется клиенту, от которого система также ждет согласия и оплаты заказа. Таким образом, на различных стадиях появляется состояние ожидания.

Процессные языки и существуют для того, чтобы описывать такие большие сценарии. Однако очень важно различать сценарии, выполняемые в одной и той же системе («оркестрация»), и сценарии, которые описывают протоколы обмена между составными системами («хореография»). GOP ориентируются только на процессные языки, реализующие сценарий «оркестрации». Однако можно организовывать обмен сообщениями между гетерогенными системами за счет интеграции BPEL и jDPL языков (jDPL поддерживает работу с BPEL).

Чтобы описать вышеприведенный сценарий полностью, нужна система, которая позволит работать с состоянием ожидания. В большинстве приложений данные на время ожидания должны быть сохранены, поэтому простого блокирования потоков недостаточно. Опытные Java-программисты могли бы вспомнить методы Object.wait() и Ob-

ject.notify(), которые позволяют симулировать состояние ожидания, но проблема в том, что эти потоки не сохраняются. Об этом уже должен заботиться программист.

Графическое представление

В некоторых областях разработки ПО можно получить значительный выигрыш во времени и ресурсах, затрачиваемых на процесс разработки, используя графическое представление бизнес-логики. Управление бизнес-процессами (Business Process Management, BPM) – одна из самых очевидных таких областей применения. Коммуникация между бизнес-аналитиком и программистом улучшается, если используется некий общий язык для описания программируемых процессов в виде графического представления.

Однако существует большое различие между диаграммами UML и диаграммами, написанными для GOP. Дело в том, что UML описывает статические модели, или просто структуры данных ООП, и созданная в нем диаграмма никак не может быть связана с реальными действиями системы. Графическое представление также выглядит как потерянное звено в ООП, потому что прямого отношения между ОО программой и ее графическим представлением нет.

В GOP описание графа – центральный и необходимый компонент, так же как XML файл, который его описывает. С тех пор, как диаграмма является внутренней частью программы, они всегда синхронизованы, и нет необходимости в ручном переводе графических требований в дизайн приложения. Программа строится на основе графовой модели.

Грамматика GOP

GOP-программа строится на основе интерпретации графа в режиме реального времени (при инициализации, и далее все инструкции сохраняются в источник данных datasource, например, базу данных), т.е. во время выполнения программы. В этом разделе будет показано, как исполнение графа может быть реализовано в среде выполнения ООП. Под «исполнением графа» в данной статье подразумевается выполнение инструкций, относящихся ко всей диаграмме (или части ее), описанной в виде узлов, связей между ними, и действиями, которые могут быть сопоставлены с различными событиями, о которых пойдет речь дальше. «Исполнение графа» может быть представлено как движение маркера (token) по линиям (траектории) выполнения процесса. Для тех, кто знаком с шаблонами проектирования, это может выглядеть, как комбинация Шаблона команд (command pattern) и Шаблона цепочки ответственности (chain of responsibility pattern).

Прежде всего, структура графа представляется классами Node и Transition (рис. 2). Класс Transition (или переход) имеет направление, поэтому граф может иметь как входящий переход, так и исходящий. В свою очередь, Transition обязан иметь и вход, и выход, в то время как Node может иметь либо вход, либо выход, либо оба (рис. 2).

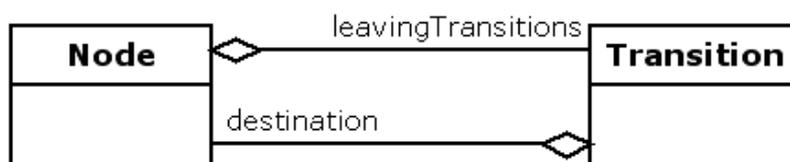


Рис. 2. Схема классов Node (узел графа) и Transition (ребро графа, переход)

Класс Node – это класс, являющийся образом «графа» в программной модели, и имеет метод execute(). Подклассы этого класса переписывают этот метод, чтобы реализовать собственное, специфическое поведение.

Модель исполнения, которая определена для графа, похожа на конечный автомат или диаграмму состояния в UML. В действительности GOP может быть использовано, чтобы реализовать любой тип поведения.

Исполнение графа (маркер, token) представляется в виде класса Execution и имеет ссылку (связь) на текущий узел (Node) (рис. 3).

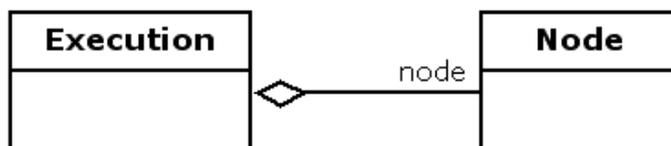


Рис. 3. Execution класс

Переходы (transitions) могут передавать выполнение от одного узла (источник, source) другому (пункт назначения, destination) с помощью метода take() (рис. 4).



Рис. 4. Transition take() метод

Когда «исполнение» (маркер) «прибывает» к узлу, этот узел запускает на выполнение метод execute() (рис. 5), который также ответственен за продолжение хода выполнения. «Продолжение исполнения» означает, что узел может передать маркер выполнения следующему узлу через исходящий переход.

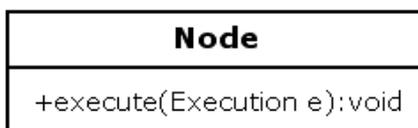


Рис. 5. Node execute() метод

Если в текущем узле метод execute() не продолжает выполнение процесса (не перемещает маркер выполнения к следующему узлу), то такое поведение соответствует состоянию ожидания. Также, когда создается новый маркер выполнения процесса (процесс инициализируется в некотором стартовом узле), процесс переходит в состояние ожидания, пока не произойдет некоторое событие.

Когда событие передается маркеру исполнения, это дает начало движению процесса к следующему узлу. Если данное событие связано с исходящим переходом (Transition), маркер выполнения захватывает этот переход. Выполнение продолжается дальше, пока не войдет в следующий узел, и снова переходит с состояние ожидания (рис. 6).



Рис. 6. Execution event() метод

И состояние графа (узла), и состояние маркера выполнения процесса могут быть сохранены: например, в реляционной базе данных, посредством ORM или объекты мо-

гут быть сериализованы в файл. В jPDL об этом разработчику беспокоиться не нужно. По умолчанию используется широко распространенная ORM-библиотека библиотеки Hibernate [8].

Процесный язык

Процесный язык в данном случае представляет собой всего лишь набор реализаций Node класса. Каждый подкласс реализует определенное поведение и соответствует определенной процессной модели.

На рис. 7 показаны 4 базовых типа узлов: start state (узел начала процесса), decision (узел принятия решения), task (узел, создающий задачу) и end state (узел конца выполнения).

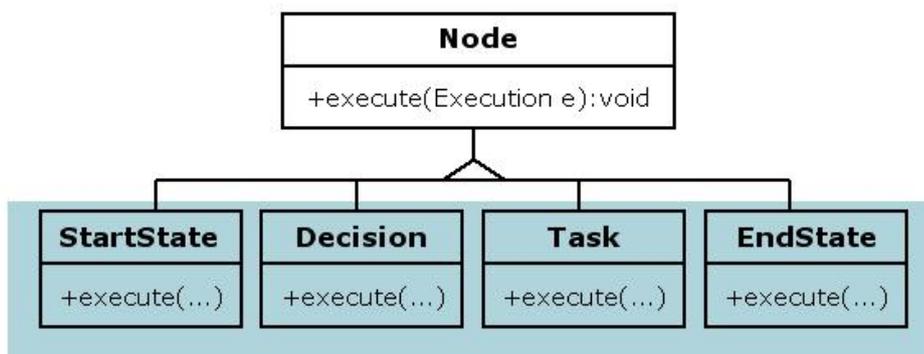


Рис. 7. Подклассы базового класса Node, реализующие разные типы узлов в процесном языке

Используя приведенные выше узлы и связывая их между собой переходами, можно построить некоторую графовую модель процесса (рис. 8).

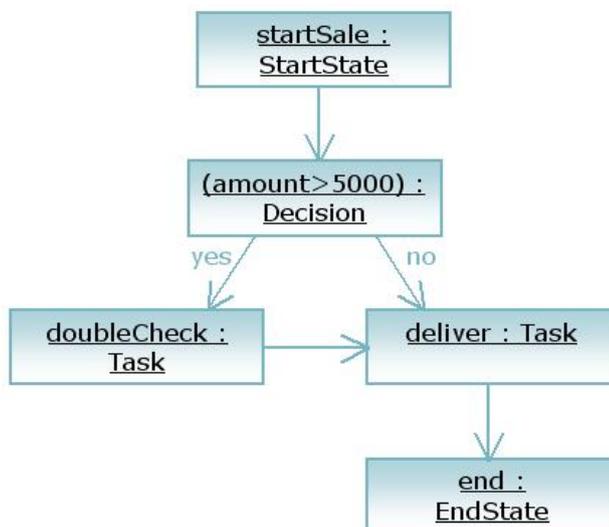


Рис. 8. Пример графовой модели процесса

Когда создается новый экземпляр этого процесса (рис. 9), маркер выполнения позиционируется в начальном узле (start state). До тех пор, пока не будет получено некоторое событие, маркер будет оставаться в том же узле (start state).

Теперь посмотрим, что произойдет, когда будет получено событие. В данном случае вызывается метод event() объекта Execution. Event(). Метод находит исходящий переход «по умолчанию» и передает Execution через переход Transition, вызывая метод take() и передавая себя в качестве параметра.

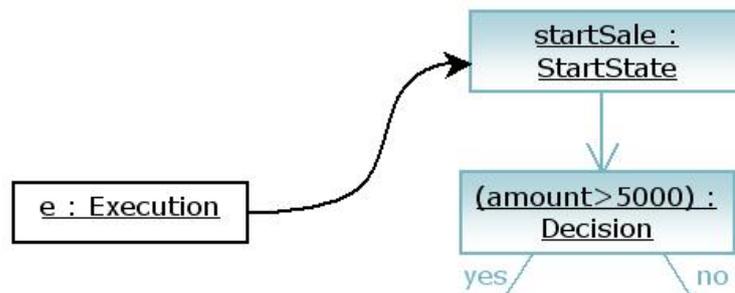


Рис. 9. Создание нового маркера выполнения процесса

Маркер перемещается к узлу Decision, и опять вызывается его метод execute(). Предположим, что этот метод производит некоторые вычисления и решает продолжать выполнение по ветви yes. Посылается 'yes'-event маркеру выполнения, и он переходит к узлу 'doubleCheck'. Теперь предположим, что реализация doubleCheck Task узла создает новую задачу и добавляет ее в список задач проверяющего, и до тех пор, пока проверяющий не отправит подтверждение, процесс будет находиться в состоянии ожидания (рис. 10) – процесс добавления к списку в случае jPDL автоматизирован, т.е. в логике приложения указывается ID соответствующего пользователя.

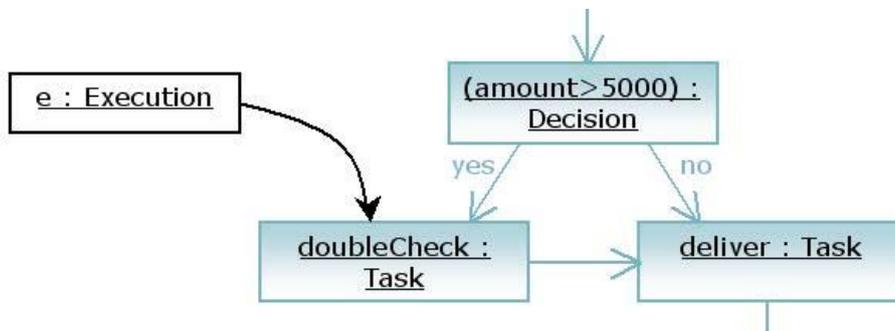


Рис. 10. Маркер выполнения ожидает в узле 'doubleCheck'

А когда придет подтверждение, маркер перейдет к следующему узлу, выполнять следующие инструкции.

Действия

В некоторых областях применения этой техники существует потребность добавлять программную логику без введения новых типов узлов для них. В BPM системах это особенно важно. Бизнес-аналитик отвечает за графическую модель, разработчик – за программную логику. Неприемлемо, если разработчик будет изменять графическую модель из-за того, что необходимо будет включить какие-то технические детали реализации, которые бизнес-аналитику неинтересны.

Для этого существуют так называемые действия (Actions) – это такие же команды с методом execute(). Действия связаны с событиями.

Существует два базовых события, вызываемые Node классом во время выполнения процесса – node-leave (выход из узла) and node-enter (вход в узел). Эти события делают возможным включать любую программную логику в зависимости от ситуации: при входе в узел или при выходе из него. Аналогично дело обстоит с переходами (см. рис. 11).

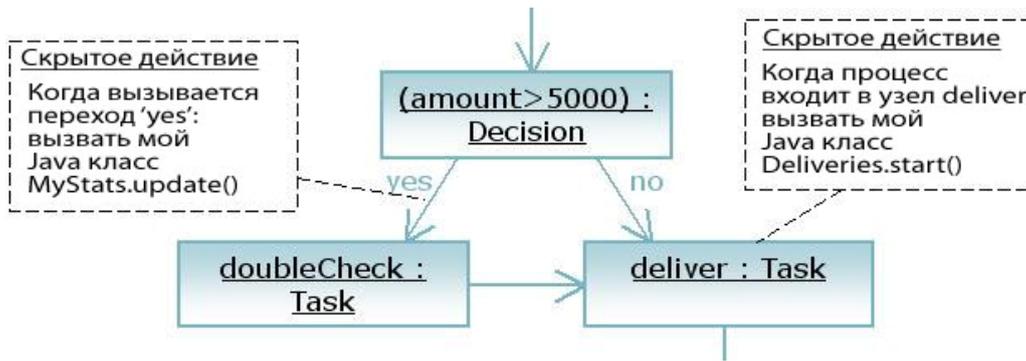


Рис. 11. Действия, которые обычно скрыты в графическом представлении

В листинге 1 показан пример класса Действия:

Листинг 1: Пример кода для Action класса

```

package com.ifmo;

import org.jbpm.graph.def.ActionHandler;
import org.jbpm.graph.exe.ExecutionContext;
public class MyStats implements ActionHandler {
    public void execute(ExecutionContext context) throws Exception {
        //некоторые действия, например вызов некоего метода, который
        //выполняет сохранение данных в базу данных, которые
        //пришли с веб-формы
    }
}

```

Одно обязательное условие – класс должен реализовывать интерфейс ActionHandler, а также метод `public void execute(ExecutionContext context)`.

В листинге 2 показан итоговый XML-файл описания процесса с описанием классов, которые будут вызываться автоматически при определенных событиях.

Листинг 2: Пример xml-файла описания процесса

```

1 <process-definition xmlns="urn:jbpm.org:jpd1-3.2" name="simple">
2   <start-state name="StartSale">
3     <transition name="" to="(amount>5000)"></transition>
4   </start-state>
5   <decision name="(amount>5000)">
6     <transition name="yes" to="doubleCheck">
7       <action name="update" class="com.ifmo.MyStats"></action>
8     </transition>
9     <transition name="no" to="deliver"></transition>
10  </decision>
11  <task-node name="doubleCheck">
12    <transition name="" to="deliver"></transition>
13  </task-node>
14  <task-node name="deliver">
15    <event type="node-enter">
16      <action name="deliverStart"
17 class="com.ifmo.Deliveries"></action>
18    </event>
19    <transition name="" to="end1"></transition>
20  </task-node>
21  <end-state name="end1"></end-state>
22 </process-definition>

```

Верхний `process-definition` xml-схемы содержит полное описание процесса. Разумеется, в одной xml-схеме такой элемент должен быть единственным. В качестве атрибута `name` принимается имя процесса (строка 1). Это имя должно быть уникально для всех процессов в системе. Процесс можно найти по его имени с помощью метода `findProcessDefinition()` класса `org.jbpm.db.GraphSession`. Узел `start-state` определяется в строке 2, также имеет имя, определяемое атрибутом `name`, также может быть только один для одного процесса (как и `end-state` в строке 21) и соответствует узлу Start State, описанному выше (рис. 8). В строке 3 определяем переход, который не имеет имени, а также связку узла StartState со следующим с помощью атрибута `to`. В строке 5 определяется узел `decision` (таких узлов может быть сколько угодно, однако имя каждого узла должно быть уникально). В строках 6 и 9 имеется два выхода `transition` из этого узла: `yes` и `no` (рис. 10). С переходом `yes` связывается вызов события `<action name="update" class="com.ifmo.MyStats">`, где указывается его имя и класс, с которым оно ассоциируется. Во время работы в этой точке будет вызван метод `execute()` этого класса, как описывалось ранее. В строке 15 событие связано с узлом, а именно с действием `type="node-enter"`, т.е. вход в узел. Аналогично будет вызван метод `execute()` класса `com.ifmo.Deliveries`. Еще один тип узла используется в строке 14, `task-node`, который создает задачу для пользователя и ждет, пока задача не будет завершена. Вся эта работа скрыта от создателя диаграммы. Хорошей практикой является вынесение части данных, используемые в логике, в этот файл с помощью переменных и правил (в данной статье не рассматривается) – например, имени пользователя, которому должна быть передана задача. Таких узлов также может быть сколько угодно. Выполнение процесса завершается, когда маркер выполнения дойдет до узла `end-state` (строка 21).

Параллельное выполнение

Допустим, разрабатывается процесс оформления покупки на основе GOP для workflow (потока работ). После того, как клиент отправил заказ, некоторые действия по оплате и доставке товара должны быть произведены параллельно. Для этого предусмотрены Fork и Join узлы, которые начинают параллельное выполнение и завершают его, соответственно (рис. 12). Задача может быть решена на основе многопоточного выполнения процессов, и пока обе ветви не будут завершены, дальнейшее выполнение процесса будет невозможно.

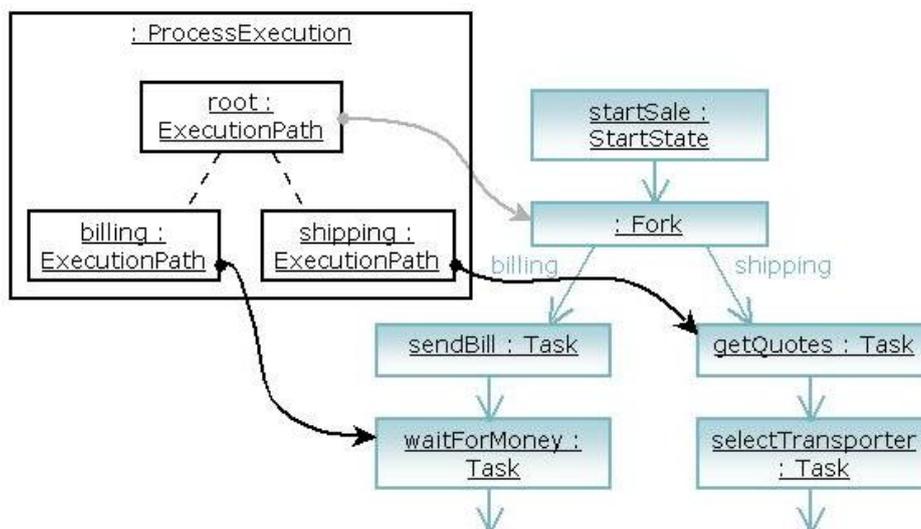


Рис. 12. Параллельное выполнение процессов

GOP на основе jPDL не использует возможности для анализа, прогнозирования, автоматического создания схем бизнес-процессов, однако, используя методы порождающих грамматик [9], можно на основе базовой модели jPDL создавать ее расширения самостоятельно, включая интеллектуальные механизмы для поддержки и обслуживания системы управления бизнес-процессами [10].

Достаточно большое количество шаблонов проектирования процессного программирования описано на сайте [11]. Помимо jPDL, существуют и другие реализации GOP, как свободно распространяемые, например, язык YAWL [12], так и для платных продуктов компаний Microsoft, Oracle и др.

Однако, по мнению автора, преимуществом jPDL языка является не только тот факт, что он относится к классу open source (свободно распространяемых) продуктов, но и то, что он имеет возможность реализовать программную логику – связку с ООП на Java, имеет виртуальную машину исполнения процессов. jPDL, может работать и как настольное приложение, и как серверное приложение, так как является стандартным java архивом (jar) и легко интегрируется с самым распространенным бесплатным сервером приложений уровня предприятия jBoss Application Server [13]. Также существует плагин для IDE Eclipse, который позволяет автоматически генерировать веб-формы, процессы, а также имеет графический редактор.

jPDL, кроме вышеперечисленных возможностей, имеет поддержку асинхронных действий, таймеров, расписаний, вложенных процессов, поддерживает работу с правилами jBoss Drools, транзакциями, автоматической отправкой сообщений email, логированием, системой безопасности, интеграцию со сторонними базами данных, уже содержащих описание ролей и пользователей системы, и многое другое, что не рассматривалось в данной статье, но описано в документации по jPDL [1].

Заключение

В статье описан новый подход к программированию гибко настраиваемых и легко изменяемых приложений, обрабатывающих какие-либо процессы с большим количеством условий и постоянно меняющимися требованиями – граф-ориентированное программирование. Основное отличие GOP от ООП заключается в том, что он реализует состояние ожидания и графическое представление программной логики, которые являются пропущенным звеном в ООП, однако сегодня имеют все большее и большее значение для построения больших автоматизированных систем управления бизнес-процессами предприятия.

К сожалению, формат статьи не позволяет осветить все аспекты программирования на основе графовой модели, однако это и не нужно. Были показаны основные базовые возможности, которые реализуют большинство процессных языков, в том числе jPDL. Как в свое время язык программирования Java стал значительным шагом в объектно-ориентированном программировании, так и GOP позволит достигнуть новых возможностей как для разработчиков ПО, так и для бизнеса.

Литература

1. JBoss.org. Официальный сайт. – Режим доступа: <http://www.jboss.org/jbossjbpm/jpdl/>
2. Martin Fowler. Официальный сайт. – Режим доступа: <http://www.martinfowler.com/bliki/DomainSpecificLanguage.html>
3. Биберштейн Н., Боуз С., Фиаммант М., Джонс К. Компас в мире сервис-ориентированной архитектуры (SOA). – М.: Кудиц Пресс, 2008. – 228 с.

4. JBoss Seam. Официальный сайт. – Режим доступа:
<http://www.jboss.com/products/seam>
5. G. Kiczales, J. Lamping, A. Mendhekar, Aspect-Oriented Programming // European Conference on Object-Oriented Programming (ECOOP), Finland. Springer-Verlag LNCS 1241. June 1997.
6. Martin Fowler. Официальный сайт. – Режим доступа:
<http://www.martinfowler.com/articles/injection.html>
7. Martin Fowler. Официальный сайт. – Режим доступа:
<http://www.martinfowler.com/articles/languageWorkbench.html>
8. Hibernate. Официальный сайт. – Режим доступа: <http://www.hibernate.org/>
9. Калянов Г.Н. Моделирование, анализ, реорганизация и автоматизация бизнес-процессов. – М.: Финансы и статистика, 2006. – 238 с.
10. Власов В.В. Автоматизированная интеллектуальная система поддержки распределенной информационной архитектуры предприятия на основе бизнес-процессного управления (BPM). / Материалы Всероссийского форума студентов, аспирантов и молодых ученых «Наука и инновации в технических университетах». – СПб: Изд-во Политехнического университета, 2007. – С. 54
11. Workflow Patterns home page. – Режим доступа:
<http://www.workflowpatterns.com/>
12. YAWL: Yet Another Workflow Language. – Режим доступа: [http://_www.yawl-system.com/](http://www.yawl-system.com/)
13. JBoss Seam. Официальный сайт. – Режим доступа: <http://www.jboss.org/jbossas/>

НЕЙРОСЕТЕВОЙ МЕТОД ОПТИМИЗАЦИИ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ СЛОЖНЫХ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

Д.В. Соловьев

Научный руководитель – д.т.н., профессор Ю.А. Гатчин,
к.т.н., доцент И.Б. Бондаренко

Данная статья посвящена нейросетевому подходу в оптимизации математических моделей сложных технологических процессов. Вводятся понятия сложного технологического процесса, требования к моделям сложных систем, кратко рассматривается технологический процесс вытяжки оптического волокна и выводится его математическая модель. Показана сложность контроля качества оптоволокна и необходимость применения адаптивных методов оптимизации на основе технологий искусственного интеллекта.

Введение

Введем понятие и выделим особенности сложного технологического процесса (ТП). Особенности сложных ТП оптического производства являются:

- малая информативность ТП из-за сложности или невозможности контроля выходных параметров;
- сложность физико-химических явлений, протекающих в ходе ТП, исключая построение аналитических математических моделей ТП;
- нестационарность ТП, являющаяся следствием физики ТП и изменений параметров, характеризующих непостоянство свойств технологического оборудования во времени;
- распределенность параметров, которая возникает из-за наличия движущихся потоков оптических материалов, а контроль параметров происходит в локальных областях или косвенными путями;
- длительность и многостадийность процесса изготовления оптических материалов;
- наличие множества перекрестных связей между отдельными каналами управления, приводящих к взаимосвязи управляющих воздействий от устройства управления (УУ).

Эти особенности сложных ТП для производства оптических материалов требуют качественной стабилизации всех параметров и наличие базы знаний по всем параметрам. Продолжительность сложных ТП может составлять несколько суток, поэтому за стабилизацию процессов отвечает автоматизированная система. Таким образом, сложные ТП являются нестационарными, нелинейными с большой степенью многосвязности и распределенностью параметров [1, 2].

Требования к моделям сложных систем

Теперь следует выдвинуть требования к математическим моделям сложных систем.

Во-первых, для оценки статистической совместимости представленная модель должна обеспечивать имитацию условий проведения натуральных испытаний сложной системы.

Во-вторых, для осуществления анализа чувствительности в состав модели должны быть включены блоки, позволяющие выбирать условия моделирования и запоминать промежуточные результаты так, чтобы расчет функций чувствительности или производных вероятностных характеристик по различным параметрам модели можно было осуществлять по возможно меньшему числу реализаций на модели.

В-третьих, для реализации основной процедуры, предусматривающей оценку показателей эффективности, в составе модели должны быть разработаны блоки, предна-

значенные для выбора условий испытаний на модели в соответствии с принятым методом математического моделирования.

От качества математических моделей, т.е. их адекватности и содержательности, во многих степени зависит практическая ценность результатов оптимизации. Необходимо определить соответствие модели реальному объекту, чтобы оценить погрешность расчета критериев качества [3–5].

Имитационная модель ТП

Проектирование и разработка систем управления сложными ТП требует учета значительного числа факторов, связанных с достаточно сложной динамикой объекта и процесса управления. Это приводит к громоздким моделям, которые трудно сформулировать, чтобы получить готовое решение. Значит, в этом случае целесообразно перейти на имитационные модели, в которых программно реализуются основные моменты динамики функционирования объекта и СУ.

Программная имитационная модель разбивается на модули, отвечающие за отдельные элементы поведения сложной системы. При использовании имитационного моделирования выделим основные этапы исследования разрабатываемой системы:

1. определение и описание системы. Осуществляется динамическое и статическое представление системы с учетом знаний, доступных на данном этапе;
2. формулирование модели. Переход от описания системы к структурной схеме моделирования и конструированию модели. Все элементы описания должны быть интерпретированы отдельными элементами из построенной системы моделирования;
3. составление программной модели. Осуществляется переход к подробным структурным схемам и программированию;
4. подготовка исходных данных. Отображение данных с учетом принятых допущений и упрощений моделирования;
5. отладка модели;
6. проверка адекватности модели – оценивается тем, насколько основные моменты системы правильно интерпретированы аналитиком, или проверкой исходных предположений, или проверкой преобразования информации от входа к выходу;
7. планирование расчетов на модели. Планирование эксперимента позволяет акцентировать внимание на наиболее существенных параметрах исследуемой системы;
8. расчеты и вывод результатов в виде графиков и таблиц;
9. анализ полученных результатов, позволяющий проектировщику принимать рациональные решения и переходить к новым вариантам проектируемой системы при необходимости;
10. документирование результатов моделирования [6, 7].

ТП вытяжки оптического волокна

В нашем случае оптимизации подлежит ТП получения оптического волокна на MCVF-установке (рис.1.). Метод MCVF (Modified chemical vapor deposition – модифицированного химического парофазного осаждения) основан на парофазном осаждении высокочистых материалов на внутреннюю поверхность кварцевой трубы, находящейся в состоянии вращения и обогреваемой горелкой.

Основными стадиями процесса являются:

1. приготовление смеси парогазовых компонентов;
2. химическая реакция окисления хлоридов с одновременным осаждением их и остекловыванием на внутренней поверхности трубы;
3. коллапсирование трубы с осажденными слоями и образование преформы;

4. вытягивание оптоволокна из образовавшейся заготовки.

Исходными компонентами для получения синтетического стекла в MCVD-процессе являются газообразные галогениды кремния, германия, фосфора и бора.

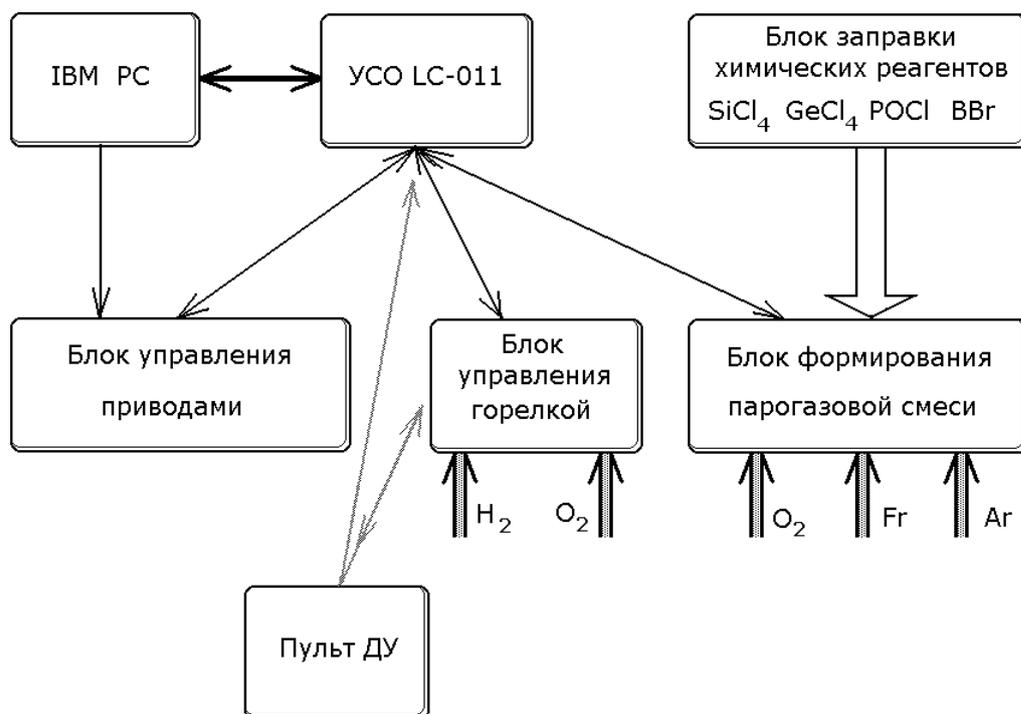


Рис. 1. Схема установки для вытяжки оптического волокна

После получения заготовки-преформы переходят к процессу формирования тонкого волокна. Для этого конец преформы нагревается в печи до пластичного состояния. При этом тянущее устройство вытягивает высоковязкий расплав в тонкую нить требуемого диаметра. Такой ТП становится возможным из-за наличия у вытягиваемого вещества достаточно широкого температурного интервала вязко-пластичного состояния или зависимости вязкости от температуры.

В исследуемом ТП наиболее важными являются температурные зависимости поверхностного натяжения и вязкости материала, которые сильно отличаются друг от друга:

$$\eta = \eta_0 e^{-\frac{E_a}{RT}}, \quad (1)$$

где: η_0 – вязкость в центре волокна; E_a – энергия активации вязкого течения (для кварцевого стекла энергия активации составляет 600 кДж/моль); T – абсолютная температура; R – газовая постоянная.

Температурные границы области вытягивания взаимосвязаны с величиной усилия вытяжки. При температуре ниже верхней границы отжига, где невозможна пластическая деформация материала, а также, если скорость вытягивания слишком велика и расплав начинает проявлять упруго-пластичные свойства и разрывается, процесс вытягивания становится невозможным. В свою очередь, скорость вытягивания волокна определяется усилием вытягивания. При сопоставлении процессов упругой и пластической деформации твердых тел и вязкости упругих тел в условиях высокотемпературного нагружения вытекает зависимость между временем деформации, усилием и вязкостью [8]:

$$\Delta l = \frac{Fl\tau}{3\pi R^2 \eta}, \quad (2)$$

где: Δl – удлинение образца; F – сила натяжения; l – длина образца; τ – время деформации; R – радиус образца. Учитывая квазистационарность процесса вытягивания волокна, сделано следующее преобразование зависимости (2):

$$V_g = \frac{\Delta l}{\tau}; A = \frac{l}{3\pi R^2} \Rightarrow V_g \eta = AF, \quad (3)$$

где A – постоянная для данного типа волокна, в которую входят геометрические размеры. Из соотношений (1) и (3) имеем:

$$V_g = \frac{AF}{\eta_0} e^{\frac{E_g}{RT}}. \quad (4)$$

Полученная зависимость (4) будет исследована в работе с помощью нейросетевого метода оптимизации.

Показатели качества оптического волокна

Для контроля сложного ТП вытягивания оптического волокна исследуются следующие показатели качества:

- геометрия оптоволокна – отклонение диаметра от требуемого значения, цилиндричность;
- профиль показателя преломления;
- затухание в оптоволокне;
- прочность волокна.

На качество оптоволокна влияет как множество входных параметров, в том числе качество заготовки-преформы, так и множество случайных воздействий, действующих извне. Поэтому учет всех воздействующих факторов, приводящих к ухудшению качества получаемого оптического волокна, является сложной задачей. Поставленную задачу предполагается решить с помощью математического аппарата искусственных нейронных сетей.

Нейросетевой подход

Анализ работ, связанных с использованием нейронных сетей для решения физико-математических задач, показал, что нейросетевой подход имеет преимущества перед традиционными математическими методами в трех случаях:

- задача в силу конкретных особенностей не поддается адекватной формализации, поскольку содержит элементы неопределенности, не формализуемые традиционными математическими методами;
- задача формализуема, но в настоящее время отсутствует аппарат для ее решения;
- решение задачи не удовлетворяет требованиям получения решений по времени, размеру, весу, энергопотреблению и др. (в данной ситуации приходится либо производить упрощение алгоритмов, что снижает качество решений, либо применять соответствующий нейросетевой подход при условии, что он обеспечит нужное качество решения задачи).

Теория искусственных нейронных сетей (ИНС) вытекает из множества дисциплин, включая физиологию, математику, нейробионику, физику и биологию, т.е. ИНС-технология – результат работы многих наук по одному направлению – созданию интеллектуальных систем. Нейроны являются основой ИНС. Искусственные нейроны представляют собой математическую модель биологических нейронов. Нейроны обменива-

ются сигналами друг с другом, образуя нейронную сеть. Эта сеть образуется с помощью многочисленных отростков нейронов: аксоны служат для передачи информации другим нейронам, а дендриты позволяют принимать сигналы. Тело нейрона суммирует прошедшие импульсы, одни из которых стремятся возбудить нейрон, другие – этому воспрепятствовать. Если суммарное возбуждение превысит определенный порог, нейрон переходит в возбужденное состояние, посылая по аксону сигнал другим нервным клеткам. В основе построения искусственных нейронных сетей лежит упрощенное представление о деятельности биологических нейронов. Схема искусственного нейрона представлена на рис. 2.

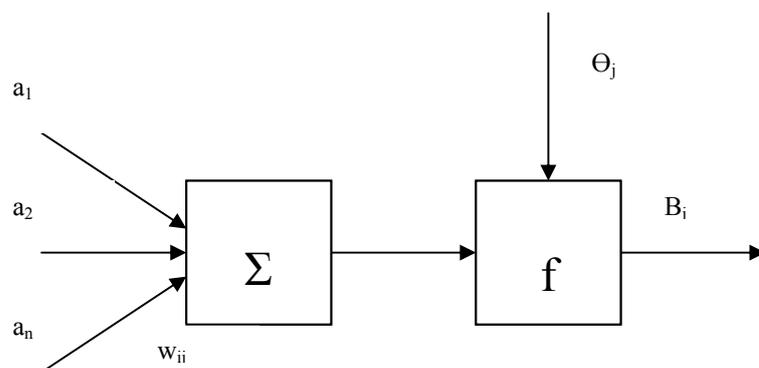


Рис. 2. Искусственный нейрон

Узел нейронной сети получает сумму взвешенных входных сигналов, сравнивает ее с порогом срабатывания узла и осуществляет над полученным результатом заданное преобразование. Сигналы поступают на входы элемента, образуя некоторый вектор $A = [a_1, a_2, \dots, a_n]$; их значения, перемноженные на весовые коэффициенты, суммируются. Весовые коэффициенты образуют весовой вектор $W_j = [w_{1j}, w_{2j}, \dots, w_{nj}]$, где коэффициенты w соответствуют связи от i -го нейрона к j -му. Результат суммирования поступает на пороговый элемент, который характеризуется величиной порога Θ_j и функцией активации B_j сигнала на выходе:

$$B_j = f(\sum a_i w_{ij} - \Theta_j). \quad (5)$$

С помощью аппарата ИНС учет вероятностного характера входных воздействий ТП будет осуществлен в виде сигналов узла нейронной сети. С учетом сложного характера воздействий, структура искусственной нейросети будет многослойной. Оптимизируемым параметром будет являться качество оптического волокна, т.е. функционал вида

$$Q(X) = F(Q_i(X)), \quad Q_i(X) = Q_i(x_1, x_2, x_3, \dots, x_n), \quad (6)$$

где $Q_i(X)$ – локальные показатели качества, $Q(X)$ – глобальный показатель качества.

Таким образом, задача получения максимального качества математически сведется к отысканию оптимального значения глобального показателя качества $Q(X)$. Из-за сложности или невозможности аналитического построения зависимости глобального показателя от локальных и из-за сложности учета влияния параметров ТП на $Q(X)$ необходимо использовать методы оптимизации, основанные на адаптивных технологиях. Поэтому использование ИНС для решения задачи оптимизации модели ТП вытягивания оптоволокна является актуальным [9–13].

Выводы

Резюмируя все вышесказанное, можно сформулировать следующие выводы:

1. технологический процесс вытяжки оптического волокна относится к сложным ТП;
2. для проведения оптимизации ТП его модель должна быть адекватна и содержательна;

3. для упрощения расчетов необходим процесс перехода на имитационные модели ТП;
4. получена математическая модель ТП вытяжки оптического волокна;
5. качество оптоволокна достаточно сложно контролировать в процессе его производства;
6. для оптимизации качества предложено использовать нейросетевой подход.

Литература

1. Бондаренко И.Б., Андреев А.К., Гатчин Ю.А. Интегрированная система автоматизированного производства оптических материалов. / Тезисы докладов научн.-технич. международной конференции «Прикладная оптика-96». – Санкт-Петербург, 1996. – С.109.
2. Бондаренко И.Б., Гатчин Ю.А. Методы автоматизированного проектирования сложных технологических систем по производству оптических материалов. / Тезисы докладов 29 научно-технической конференции «Проектирование и технология элементов компьютерных систем». – СПбГИТМО, 1997. – С. 6–7.
3. Алексеев А.В. Интеллектуальные системы принятия проектных решений. – Рига, 1997.
4. Гатчин Ю.А., Коробейников А.Г. Проектирование интегрированных автоматизированных технологических комплексов. – СПб: СПб ГИТМО (ТУ), 2000.
5. Бондаренко И.Б. Методы оптимального проектирования сложных технологических систем / Тезисы докладов 31 межвузовской научно-технической конференции профессорско-преподавательского состава. – СПбГИТМО, 1999. – С. 86.
6. Бондаренко И.Б., Гатчин Ю.А. Оптимизация выбора проектных решений. /Тезисы докладов 30 межвузовского научно-технической семинара с международным участием «Автоматизация проектирования, технология элементов и узлов компьютерных систем». – СПбГИТМО, 1998. – С. 7–8.
7. Бондаренко И.Б., Гатчин Ю.А. Оптимизация проектных решений в САПР автоматизированных технологических комплексов. // Научно-технический вестник СПбГИТМО (ТУ). – 2002. – Выпуск 6. Информационные вычислительные и управляемые системы. – С. 127–135.
8. Соломин Н.Д. Жаростойкость материалов и деталей под нагрузкой. – М.: Стройиздат, 1969.
9. Научная сессия МИФИ-2003, V Всероссийская научно-техническая конференция «Нейроинформатика-2003», лекции по нейроинформатике. Часть 1. – М.: МИФИ, 2003. – 188 с.
10. Вороновский Г.К., Махотило К.В., Петрашев С.Н., Сергеев С.А. Генетические алгоритмы, искусственные нейронные сети и проблемы виртуальной реальности. – Х.: ОСНОВА, 1997. – 112 с.
11. Заенцев И.В. Нейронные сети: основные модели. / Учебное пособие. – Воронеж, 1999.
12. Дианов Р.С. Оптимизация технологического процесса разработки газоносного пласта с применением генетических алгоритмов и нейронных сетей: Дис. ... канд. техн. наук: 05.13.06 Астрахань, 2004 167 с.
13. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика. / Перевод Ю.А. Зувев, В.А. Точенов. – 1992.

ИСПОЛЬЗОВАНИЕ САПР ПРИ СОЗДАНИИ «УМНЫХ ДОМОВ»

П.В. Федосов

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье рассматриваются возможности «умных домов» и использование САПР при их проектировании. Так же приведено сравнение возможностей различных САПР в применении к данной задаче. Статья будет полезна тем, кто хочет подробнее узнать о процессе разработки систем «умный дом».

Введение

С каждым годом нас окружает все больше сложной электронной аппаратуры. Казалось бы, еще лет 15 назад даже пара современных телевизоров была далеко не в каждой квартире. Сейчас уже мало кого удивит наличие трех-четырех телевизоров, CD/DVD, ресиверов спутникового телевидения к ним, и это еще не считая компьютерную и бытовую технику. Думаю, многие уже не раз сталкивались с проблемой поиска нужного пульта для одного из компонентов электронного зоопарка или с многочасовыми бдениями над инструкциями в поисках пути правильного подключения различных электронных компонентов друг к другу.

Системы класса «умный дом» как раз предназначены для того, чтобы избавить пользователя от зачастую ненужной и не интересной ему технической стороны и предоставить простой и удобный интерфейс управления, в идеале, всей электроникой в доме. Однако управление мультимедийной аппаратурой – это лишь небольшая, зачастую легко разрешимая задача. Построение «умных домов», управляющих также осветительной, вентиляционной, системой кондиционирования и охранными системами, требуют гораздо более комплексного подхода, главная задача которого – минимизировать затраты на выполнение работ.

В этой статье рассмотрены применение некоторых САПР для решения задач проектирования «Умных домов» [1].

Основные возможности

Умный дом или офис – это автоматизированная система управления, которая предназначена для контроля и управления освещением, отоплением, вентиляцией, водоснабжением, безопасностью, аудио/видео аппаратурой и другими инженерными системами дома. Систему можно настраивать и модифицировать по желанию владельца. Всеми функциями системы можно управлять дистанционно с пульта или даже с мобильного телефона. Обычно умный дом позволяет интегрировать различные подсистемы, обеспечивая их слаженную работу и высокую функциональность всего комплекса [2].

Концепция умного дома подразумевает слаженную работу различных систем. Это позволяет не только исключить конфликты при их функционировании, но и обеспечить гармоничное взаимодействие. Классический пример: кондиционер не будет охлаждать помещение в то время, когда его нагревает теплый пол. Интеграция обеспечивает комфорт в жилье. Управление всем комплексом умный дом берет на себя. По одному Вашему желанию, как дирижер управляет оркестром, умный дом управляет освещением, шторами, кондиционером, теплым полом, аудио/видео аппаратурой и др. оборудованием, а Вы наслаждаетесь комфортом (см. рисунок).

Рассмотрим подробнее типовые функции:

- Управление освещением. Внутреннее освещение во всем доме, наружное освещение, динамичные световые сцены позволяющие придать разные «настроения» помещению, включение при прохождении, плавное включение ламп накаливания с учетом мощности, выключение всех устройств по одной команде. Все это позволяет рациональнее использовать ресурс осветительного оборудования.

- Управление силовыми розетками. Включение/отключение групп розеток или отдельных розеток в каждой комнате. При этом можно управлять различными бытовыми устройствами, к примеру, отключить телевизор в детской комнате, отключить питание утюга после заданного интервала времени, уходя из дома обесточить все устройства (кроме заранее определенных) по нажатию одной кнопки и т.д.



Рисунок. Пример организации взаимодействия систем

- Управление электродвигателями. Шторы, жалюзи, роллеты, автоматические ворота, шлагбаумы, гаражные ворота, насосы
- SMS-оповещение и управление. Оповещение на мобильный телефон при возникновении аварийных ситуаций – пожар, затопление водой, несанкционированное проникновение и т.п.
- Безопасность и сигнализация. Датчики движения, датчики закрытия окон и дверей обеспечивают защиту помещений от несанкционированного проникновения. К функциям безопасности также относится защита от затопления водой, пожарная сигнализация, обесточивание электроприборов при уходе, возможность отправки сообщений по SMS при аварийных ситуациях, включение сирен, имитация присутствия. Кроме того, система позволяет организовать тревожные кнопки, например, в детской комнате, у кровати больного и т.д.
- Функции «Мультирум». Система Мультирум обеспечивает матричную коммутацию аудио/видеосигналов по помещениям. Звук и изображение от различных источников (DVD и CD проигрыватели, радио и т.д.) выводится в требуемых помещениях. Управление переключением каналов, настройкой громкости выбор FM-станции по командам с тачскрина, пульта или настенных кнопок, по датчикам движения и т.д.
- Универсальный дистанционный пульт. Система, имея ИК, Bluetooth или WiFi передатчики, может управлять всеми устройствами с одного универсального пульта, сенсорной панели, кнопочных выключателей или мобильного телефона можно удаленно управлять плеерами, кондиционерами, телевизорами, CD/DVD проигрывателями во всем доме.
- Климат-контроль. Система позволяет управлять климатом как во всем доме, так и в отдельных помещениях, обеспечивая слаженную работу систем и приборов, ответственных за климат (система отопления, теплый пол, вентиляция, кондиционирование, влажность и т.д.).
- Автоматический полив. Контроль влажности почвы, управление насосами, управление по расписанию, учет погодных условий

Умный дом легко интегрируется в общую систему интеллектуального здания. Это позволяет контролировать аварийные ситуации (затопление, пожар, охрана), состояние

оборудования, температуру и другие параметры всех автоматизированных апартаментов в одном диспетчерском пункте. Из квартиры или номера в отеле можно вызывать необходимые службы. С компьютера в диспетчерской отдельные апартаменты переводятся в режим «отсутствия хозяев» или предварительно подготавливаются к приезду жильца.

Проектирование

Из сказанного очевидно, что создание системы подобного масштаба вслепую фактически невозможно. Все это делает особенно важным подготовку качественной проектной документации. Сейчас проектирование фактически любого прибора или комплекса не обходится без использования САПР. Процесс проектирования «умных домов» не является исключением. К сожалению, единого пакета схожего с продукцией компаний Solid Works или Autodesk, способного решать все задачи проектирования умных домов, на сегодняшний день не существует. Поэтому используются комбинации различных САПР для достижения желаемого результата.

Как мы уже убедились ранее, «умный дом» фактически нашпигован различными контроллерами, находящимися в каждой розетке, выключателе и т.д., которые соединены большим количеством как силовых, так и сигнальных кабелей. Также следует помнить, что реализация проектов «умных домов» делается преимущественно не по какому-то шаблону, а под требования заказчика, что порой вызывает трудности, поскольку приходится реализовывать не стандартные решения.

На данный момент для создания проектной документации наиболее часто используется программа AutoCAD от компании Autodesk. Это объясняется тем, что рассматривать схему коммуникаций между различными устройствами, а также схемы монтажа слаботочных и высоковольтного оборудования без привязки к плану этажа фактически бесполезно. Сегодня компания Autodesk является одним из мировых лидеров в области производства программного обеспечения для автоматизированного проектирования и подготовки проектной документации – с ее продуктами работают миллионы пользователей в более чем 160 странах мира.

Справедливости ради стоит отметить, что, хотя и большая часть документации по проектам создается в формате продуктов компании Autodesk, это не делает их выбор единственно возможным. Формат DWG с каждым годом становится все более популярным, и его поддержка появляется все в большем количестве сторонних программ. Наравне с AutoCAD могут быть использованы такие программы, как семейство продуктов КОМПАС компании АКСОН. Стоит также отметить, что зачастую при проектировании не требуются какие-то специфические функции, присущие только одной уникальной программе. Так, например, с успехом может применяться пакет интегрированного комплекса САПР – Solid Works. Хотя его профильной областью применения являются гибридное параметрическое моделирование, проектирование деталей, сборок и изделий с учетом специфики изготовления (листовой материал, пресс-формы и штампы, сварные конструкции), экспресс-анализ (массово-инерционные характеристики, анализ прочности и кинематики), оформление чертежей по ЕСКД, он так же может быть использован в процессе проектирования «умных домов» просто как редактор DWG. Однако в состав пакета входит модуль Routing, позволяющий проектировать трубопроводы и электрожгуты с использованием библиотек стандартных элементов, что позволяет значительно сократить время разработки.

Также зачастую применялись САПР OrCad и PCAD для проектирования плат сопряжения контроллеров различных производителей и т.п.

Сравнение

Ввиду отсутствия специализированных САПР, разработанных для проектирования «умных домов», сравним несколько пакетов САПР широкого профиля. Для удобства выбора между ними составлена таблица [3, 4]. Как видно из таблицы, нельзя отдать од-

нозначное предпочтение какому-то одному проекту, поскольку ни один из них специально не разрабатывался и не адаптировался для решения требуемых задач, мы используем их возможности не по прямому назначению или не в полной мере.

Описание	AutoCAD	Solid Works	PCAD	OrCad
Работа с планами этажей здания	+	+	–	–
Возможность совместной работы с другими направлениями проектирования	+	+	–	–
Наличие библиотеки шаблонов	+	+	+	+
Возможность расширения функционала пакета за счет подключения сторонних модулей	+	+	+	+
Генерация чертежей и проектной документации в финальном виде, не требующем доведения в других программах	+	+	–/+	+
Возможность экспорта документации в сторонние форматы	+	+	+	+
Возможность выполнения всей проектной документации без применения сторонних программ	+	+	–	+
Возможность оптимизации проектного решения средствами пакета	–	–	+	+
Возможность моделирования поведения системы	–	–	–	–

Таблица. Сравнение пакетов САПР

Заключение

В статье рассматривались возможности «Умных домов» и средства их проектирования. В настоящее время разработка и внедрение таких систем – не редкость. Таким образом, все расширяющаяся область применения таких систем требует дальнейшего развития средств разработки и автоматизации проектирования.

Существующие сейчас САПР, хоть и удовлетворяют потребности проектирования, но дальнейшее усложнение систем сделает их применение малоэффективным.

Литература

1. АВОК Стандарт 3-2003. Системы автоматизации и управления зданиями. Часть 1. Общие положения

2. Гололобов В.Н. «Умный дом» своими руками. – М.: НТ Пресс, 2007. – 416 с.
3. Кашкаров А.П. Электронные схемы для «умного дома». – М.: НТ Пресс, 2007. – 256 с.
4. Сопер М.Э. Практические советы и решения по созданию «умного дома». – М.: НТ Пресс, 2007. – 432 с.
5. Тесля Е. «Умный дом» своими руками. Строим интеллектуальную цифровую систему в своей квартире. – СПб: Питер, 2008. – 224 с.
6. Харке В. Умный дом. Объединение в сеть бытовой техники и системы коммуникаций в жилищном строительстве. – М.: Техносфера, 2006. – 290 с.

ЭЛЕМЕНТЫ ОПТИЧЕСКОЙ ПОДСИСТЕМЫ СКС

Д.А. Шилкин

Научный руководитель – Д.О. Изумрудов (ЗАО «Эврика»)

В статье рассмотрены современные оптические разъемы и соединители, которые являются неотъемлемой частью структурированной кабельной системы, а их качество в значительной степени определяет дальнейшие эксплуатационные характеристики сетей. Контроль на этапе монтажа и тестирование оптических систем позволяют добиться минимальных величин потерь и обратных отражений и достигнуть максимально заявленной скорости передачи информации в гигабитных и мультигигабитных приложениях, а при необходимости скорректировать те или иные параметры.

Введение

Современный этап развития локальных сетей характеризуется быстрым ростом их размеров, расширением объема выполняемых функций, четко наметившейся тенденцией интеграции в единое информационное пространство отделов, филиалов и офисов одной организации, часто удаленных друг от друга на десятки километров. Функционирование корпоративной информационно-вычислительной сети крупного предприятия требует передачи все большего объема разнообразных данных на все более высоких скоростях, которые требуются для современных приложений.

Традиционные системы связи на основе витой пары и коаксиального кабеля уже не в полной мере удовлетворяют современным требованиям в отношении пропускной способности, помехозащищенности, скрытности и дальности связи. Эти обстоятельства диктуют необходимость перехода к новому типу направляющей системы электромагнитных колебаний – волоконно-оптическому кабелю, а также применения специальных технических средств, предназначенных для работы с этим типом физической среды: аппаратуры волоконно-оптических систем передачи информации. Огромное значение, влияющее на качество реализации этих систем, имеют так называемые коммутационно-распределительные устройства: соединители, муфты, коннекторы, розетки, переходные разъемы и т.д.

В статье затронуты актуальные на сегодняшний день вопросы, касающиеся оптической подсистемы структурированных кабельных систем (далее СКС). Приведена система международной стандартизации оптических разъемов, описаны параметры (потери, отражение, физический контакт), а также рассмотрены основные типы соединителей, используемых в настоящий момент на российском рынке СКС.

1. Оптические разъемы

1.1. Назначение и основные требования

Одной из важнейших проблем, которую приходится решать при создании практически любой линии оптической связи, является необходимость сращивания оптических волокон (далее ОВ) друг с другом. В общем случае данная задача может быть решена двумя принципиально различными способами: с помощью разъемных или неразъемных оптических соединителей (рис. 1). Неразъемные соединители, которые иногда называются сростками, широко применяются при создании линий связи большой протяженности (например, сетей связи общего пользования масштаба города и более), а также при сборке оконечных коммутационных устройств в случае привлечения для этого технологии монтажных шнуров.

Оптические разъемы (разъемные соединители) предназначены для сращивания световодов в тех случаях, когда во время эксплуатации часто возникает потребность в изменении конфигурации трактов передачи информации или же существует повышенная опасность повреждения кабельного изделия на определенном участке тракта.



Рис. 1. Основные разновидности оптических соединителей

Оптические разъемы, которые иногда называются разъемными соединителями, предназначены для обеспечения разъемного подключения шнуровых изделий различного назначения к коммутационному оборудованию в технических помещениях различного уровня, информационным розеткам (далее ИР) рабочих мест и к сетевым устройствам [1]. В некоторых ситуациях оптический разъем (далее ОР) может выполнять функции сростка или является важным функциональным компонентом элемента специального назначения. ОР выполняет следующие основные функции:

- обеспечивает ввод ОВ в точку срачивания таким образом, чтобы радиус изгиба не уменьшился ниже предельно допустимого значения;
- защищает ОВ от внешних механических и климатических воздействий;
- фиксирует соединяемые световоды в центрирующей системе.

Перечень нормируемых эксплуатационных параметров ОР включает в себя величины максимально допустимых вносимых потерь, минимального коэффициента обратного отражения, рабочих температур и долговечности в смысле количества циклов подключения-отключения, при котором остальные характеристики разъемов не выходят за разрешенные пределы. Числовые значения этих параметров приведены в табл. 1.

Параметр	Разъемы	
	Многомодовые	Одномодовые
Номинальный диаметр оболочки волокна, мкм	125	125
Затухание, дБ	$\leq 0,75$	$\leq 0,75$
Коэффициент обратного отражения, дБ	≤ -20	≤ -35
Кол-во циклов соединения-разъединения	≥ 500	≥ 500
Рабочий диапазон температур, °С	-10...+60	-10...+60

Таблица 1. Основные характеристики оптических разъемов по ISO/IEC 11801:2002

Основной европейский стандарт EN 50173-1:2002 (Е), в отличие от американского и международного аналога, более жестко подходит к нормированию допустимого затухания: 0,5 дБ для 95% соединений и 0,75 дБ для 100% соединений.

Рабочий температурный диапазон оптических разъемов СКС согласно стандарту ISO/IEC 11801:2002 должен составлять от -10 до +60°С. На практике у большинства конструкций он колеблется от -40 до +85°С, т.е. совпадает с рабочим температурным диапазоном ОК внешней прокладки.

Отметим также относящуюся к области эксплуатации рекомендацию стандарта ISO/IEC 11801:2002 об очистке торцевых поверхностей соединяемых ОВ перед выполнением соединения ОР.

В соответствии с требованиями стандартов СКС розетки любого ОР, допускающего применение одиночных вилок, должны снабжаться фабричной или наносимой непосредственно на объекте монтажа символьной маркировкой, в виде букв А и В. Вилку с маркировкой А всегда необходимо подключать к розетке с такой же маркировкой, и наоборот, см. рис. 2. Смысл символьной маркировки вилок и розеток разъема СКС заключается в том, что она позволяет определить направление «движения» оптического сигнала. Вилка с маркировкой А всегда является источником, а розетка с такой же маркировкой – приемником, и наоборот. Аналогично на сетевом оборудовании розетка с маркировкой А является входом приемника, а с маркировкой В образует интерфейс физического уровня передатчика.

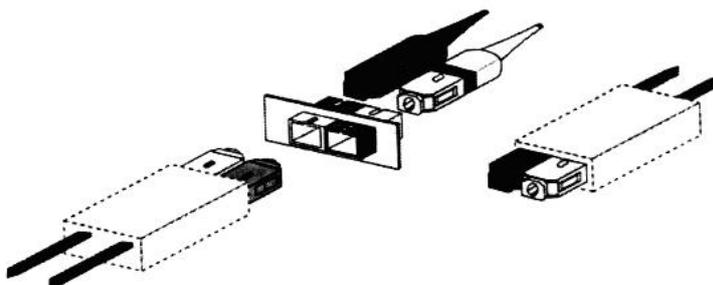


Рис. 2. Схема подключения оптического разъема

1.2. Система международной стандартизации оптических разъемов

Необходимость обеспечения совместимости компонентов ОР одного типа, изготовленных различными производящими компаниями, объективным образом привела к разработке нормативных документов, регламентирующих различные аспекты этой проблемы. В настоящее время наибольшую популярность получили два таких документа, в которых нормированы параметры наиболее распространенных типов разъемных оптических соединителей.

В Европе и азиатских странах производители элементной базы пользуются преимущественно международным стандартом IEC 61754. Структурно документ состоит из нескольких частей. Часть 1 включает в себя общие требования, предъявляемые к ОР. Остальные части содержат спецификации конкретных разновидностей разъемных соединителей для кварцевых и полимерных световодов [2].

Аналогичная стратегия принята в США: имеется базовый стандарт TIA/EIA-604 (FOCIS — от fiber optic connector intermateability), конкретные типы оптических разъемов описаны в отдельных частях этого документа, см. табл. 2.

№ части	IEC 61754	TIA/EIA-604
1	–	Bicinic
2	BFOC/2,5	BFOC/2,5
3	LSA	SC
4	SC	FC
5	MT	MTP/MPO
6	MU	FJ
7	MPO	SG-45
8	CF08	Mini-MAC
9	DS	Mini-MPO
10	Mini-MPO	LC
11	–	SCDC/SCQC
12	FC	MT-RJ

13	FC-PC	LX.5
14	–	–
15	LSH	MF
16	PN	LSH
17	–	MU
18	MT-RJ	–
19	SG	–
20	LC	–
21	SMI	–

Таблица 2. Структура стандартов серии IEC 61754 и TIA/EIA-604

2. Параметры оптических разъемов

2.1. Вносимые потери

Согласно рекомендации L.36 ITU-T, потери в оптическом разъеме определяются как

$$A = 10 \lg \left(\frac{P_{вх}}{P_{вых}} \right), \quad (1)$$

где $P_{вх}$ – мощность оптического сигнала на входе разъема; $P_{вых}$ – мощность оптического сигнала на выходе разъема; $P_{отр}$ – отраженный сигнал.

Потери в ОР определяются целым рядом причин, которые в общем виде могут быть классифицированы на следующие группы [3]:

- внутренние факторы, которые определяются допусками на геометрические размеры ОВ;
- внешние факторы, которые определяются качеством изготовления отдельных элементов разъема и его технологическими допусками;
- отражениями и рассеянием;
- загрязнениями.

К числу основных внутренних факторов, которые вызывают потери в ОР, относятся эксцентриситет и некруглость (эллиптичность) сердцевины, а также разность диаметров, числовых апертур и профилей показателей преломления сращиваемых световодов. Необходимость учета эксцентриситета и некруглости возникла на ранних стадиях развития техники оптической связи. В настоящее время в связи с достигнутым технологическим уровнем изготовления ОВ эти факторы перестали играть первостепенное значение [4].

В перечень составляющих потерь, которые вызываются внешними факторами, входят потери за счет наличия воздушного промежутка между торцами сращиваемых ОВ, радиальных и угловых смещений световодов, а также непараллельности торцевых поверхностей ОВ в разъемах. Потери этого вида обусловлены неизбежными производственными допусками на геометрические размеры отдельных деталей ОР, выполняющих центрирование сращиваемых ОВ. Потери, возникающие в случаях, когда между торцевыми поверхностями сращиваемых ОВ имеется воздушный промежуток, рассмотрены в подразделе 2.2. Потери на загрязнение возникают в процессе эксплуатации кабельной системы главным образом из-за несоблюдения правил подключения и отключения ОР. Для минимизации этой составляющей потерь стандарты требуют выполнять очистку оптически активных поверхностей соединяемых ОВ перед каждым подключением ОР.

2.2. Схема физического контакта в оптических разъемах

Основные стандарты СКС выдвигают достаточно жесткие требования по величине потерь, создаваемых разъемами оптической части структурированной проводки. Нормативные документы содержат только самые общие требования относительно кон-

структивного исполнения данных элементов кабельного тракта. Известные достоинства разъемов линзового типа при условиях, характерных для области эксплуатации СКС, оказываются не востребованными. В силу этого из соображений удобства эксплуатации, достижения высокой плотности портов на панелях коммутационного поля и т.д. предпочтительным является применение разъемов контактного типа.

Разъем контактного типа в общем случае предполагает только непосредственную стыковку соединяемых юстирующих элементов, а не самих волокон. За счет этого в собранном состоянии соединителя без применения специальных мер между ОВ потенциально может существовать воздушный промежуток. Наличие этого зазора обязательно сопровождается появлением эффекта Френелевского отражения и формирования в пространстве конусообразного пучка излучения, угол при вершине (расходимость) которого определяется числовой апертурой волокна. Данные эффекты влекут за собой увеличение потерь в разъемном соединителе.

Количественной мерой интенсивности эффекта Френеля является коэффициент отражения, равный

$$\rho = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2. \quad (2)$$

Величина френелевских потерь при $n_1 = 1,48$ на основании формулы 2 составляет $2a_F = 0,34$ дБ.

Значение второй составляющей потерь, возникающей из-за расходимости светового потока, для градиентных волокон может быть оценено соотношением

$$a_z = -10 \lg \left(1 - \frac{z}{2a} NA \right), \quad (3)$$

где z – ширина воздушного зазора между волокнами; a – радиус сердцевин. Для наиболее распространенного 50-микронного волокна с $NA = 0,2$ при $z = 5$ мкм имеем $a_z = 0,09$ дБ. Таким образом, наличие воздушного зазора увеличивает потери в соединителе на $2a_F + a_z \approx 0,4$ дБ, что является недопустимо высокой величиной.

Для устранения данной составляющей потерь оптический разъем должен быть реализован по схеме так называемого физического контакта (physical contact – PC). Данная схема основана на принудительном обеспечении различными технологическими средствами плотного прилегания друг к другу сердцевин ОВ в случае штатного подключения вилки к розетке. При этом расстояние между торцовыми поверхностями сердцевин соединяемых волокон оказывается много меньшим длины волны передаваемого излучения. Результатом является полное устранение обоих нежелательных эффектов и снижение потерь на 0,4 дБ.

Из-за конечной точности оборудования для механической обработки наконечников полного сближения торцевых поверхностей волокон практически невозможно гарантированно добиться в структурах с плоской формой торцевой поверхности, популярных в ОР, создание которых была закончено не позднее середины 80-х гг. прошлого века, см. рис. 3а. Для достижения физического контакта в современных конструкциях применяется целый ряд технических и технологических приемов, в том числе:

- использование нажимных пружин, которые при вставленных в розетку вилок прижимают торцы наконечников друг к другу с усилием примерно 8–12 Н;
- формирование у наконечников выпуклой торцевой поверхности (радиус скругления в зависимости от варианта исполнения составляет 5–25 мм), см. рис 3.

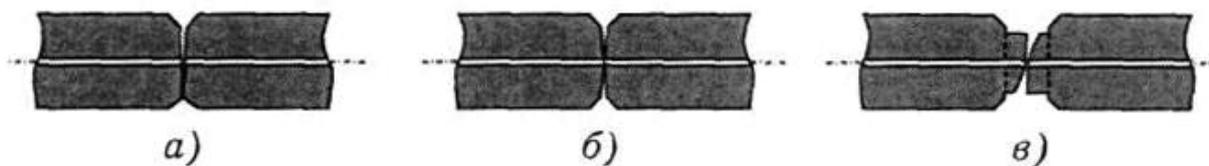


Рис. 3. Формы торцевой поверхности наконечников вилок оптических разъемов: а) плоская; б) выпуклая классов PC, SuperPC, UltraPC; в) скошенная (Angled PC)

Дополнительным положительным следствием от наличия физического контакта является то, что он позволяет в значительной степени подавить явление обратного френелевского отражения. Это является чрезвычайно важным для некоторых разновидностей одномодовых приложений.

2.3. Обратные отражения

В любом ОР между торцевыми поверхностями срачиваемых ОВ обязательно остаются воздушные зазоры большей или меньшей толщины и/или площади, вызванные неизбежными погрешностями и допусками на изготовление. В таких областях за счет наличия перехода стекло – воздух – стекло возникают френелевские отражения, которые приводят к появлению светового потока, распространяющегося в обратном направлении.

Поток обратного отражения оказывает отрицательное влияние, в первую очередь, на высокоскоростные лазерные оптические передатчики, так как, попадая обратно в резонатор излучателя, вызывает сильные искажения передаваемого сигнала [5]. В принципе, причиной возникновения обратных отражений может явиться любая неоднородность оптического тракта, однако наибольший вклад вносят ОР, где опасность их возникновения является наибольшей. На основании этого в процессе создания линий оптической связи значение обратного отражения должно контролироваться достаточно жестко. Количественной мерой интенсивности обратных отражений является коэффициент обратного отражения, который согласно рекомендации L.36 ITU-T определяется следующим образом:

$$RL = 10 \lg \left(\frac{P_{OBR}}{P_{BX}} \right), \quad (4)$$

где P_{OBR} – мощность потока обратного рассеяния; P_{BX} – мощность оптического сигнала на входе разъема.

В некоторых публикациях [6] в качестве меры обратного отражения используется эквивалентный параметр ORL (optical return loss), определяемый как

$$ORL = 10 \lg \left(\frac{P_{BX}}{P_{OBR}} \right), \quad (5)$$

т.е. $ORL = -RL$.

Нормы на коэффициент обратного отражения ОР, применяемые при построении кабельных трактов СКС, содержатся в стандарте ISO/IEC-11801:2002 и приведены в табл. 1. Указанная в этом нормативном документе величина обратных отражений для одномодовых ОР (не хуже –35 дБ) заимствована из рекомендаций ETSI. Она недостаточна для обеспечения работоспособности с заданным качеством некоторых приложений и в широкую инженерную практику внедрен ряд методов по его снижению. В зависимости от достижимого коэффициента обратного отражения одномодовые ОР традиционно делят на классы, которые приводятся в табл. 3.

Отметим также еще два обстоятельства, относящихся к области нормирования параметра обратного отражения. Во-первых, в ОР без скошенного наконечника после нескольких подключений из-за неизбежного появления царапин на торцевых поверхно-

стях сердцевин ОВ коэффициент обратного отражения устанавливается на уровне РС-разъема в независимости от уровня исходной полировки. Во-вторых, многоволоконные разъемы в одномодовом и многомодовом вариантах, как правило, имеют на 10–15 дБ меньшее значение параметра RL по сравнению с одноволоконными изделиями.

Фирма-производитель	ММ	РС	SPC	UPC	APC
Типовое значение	-20	-35	-45	-55	-65
Amphenol, США	-	-	-45	-55	-65
Diamond, Швейцария	-40	-50	-	-	-70
Huber+Suhner, Швейцария	-	-45	-50	-	-85
Molex, США	-	-	-45	-55	-70
Radiall, Франция	-20	-30	-40	-50	-60
Senko, Япония	-	-	-50	-55	-65
Telegartner, Германия	-30	-40	-	-	-60
Tyco Electronics, США	-	-50	-	-	-60

Таблица 3. Значения коэффициента обратного отражения различных классов оптических разъемов некоторых производителей

Обязательным условием минимизации обратного отражения является наличие физического контакта. Наличие физического контакта особенно важно для одномодовых разъемов. Для этих изделий наряду с мероприятиями, рассмотренными в параграфе 2.2, используют наконечники со скошенной торцевой поверхностью, см. рис. 3в.

Последний вариант является наиболее эффективным, хотя и самым сложным в технической реализации, и соответственно дорогим средством минимизации обратных отражений. Данная разновидность наконечников называется APC (от Angled physical contact, в Японии часто применяется трактовка этой аббревиатуры как Advanced Physical Polish [7]). Известно также их обозначение как PAE-, HRL- или SP-наконечники (от pre-angled endface, high return loss и slant polishing соответственно). Угол скоса выбирается несколько большим по сравнению с апертурным углом армируемого ОВ, чтобы излучение, отражаемое обратно в сердцевину, быстро высвечивалось в оболочку. В соответствии с этим условием угол скоса наконечников ОР, предназначенных для установки на волокна классов В1.1 (G.652) и В1.3 по IEC 60798-2-50, составляет примерно 8–9° (стандартные одномодовые ОВ этих классов имеют параметр NA примерно 0,12, который соответствует апертурному углу 6,8°). В тех случаях, когда вилки соединителя предназначены для установки ОВ со смещенной дисперсией, которые обладают повышенной числовой апертурой, угол наклона увеличивается до 12°. Отметим, что аппаратура, передатчики которой обладают повышенной чувствительностью к обратному отражению, работает только по одномодовым ОК. В силу этого скошенные наконечники применяются исключительно в одномодовых ОР.

3. Установка и монтаж элементов оптических разъемов

В процессе создания волоконно-оптических линий в сетях связи общего пользования сращивание ОВ и установка вилок ОР выполняются в подавляющем большинстве случаев с использованием сварочных аппаратов. Такой подход обусловлен тем, что наряду с высокой скоростью выполнения работы технология сварки обеспечивает ми-

нимальные, по сравнению с другими методами, потери в точке сращивания, а данный параметр критически важен на линиях большой протяженности, что характерно для данной области применения. Платой за эти преимущества являются значительная стоимость оборудования и повышенные требования к уровню технической подготовки персонала [8].

Элементная база для оконцевания оптических кабелей СКС может иметь две принципиально отличные друг от друга формы исполнения, каждая из которых, в свою очередь, реализуется в двух вариантах. Первая форма исполнения предполагает использование только элемента ОР (вилки или розетки), тем или иным способом устанавливаемого на ОВ. Способы реализации варьируются в зависимости от того, что поставляется на объект монтажа: набор исходных компонентов или уже частично предустановленные элементы, где ряд технологических операций с волокном был выполнен еще до выезда на объект монтажа, в стационарных заводских условиях. Обычно это наиболее ответственные операции заклейки, полировки и шлифовки, которые производятся с широким использованием автоматического и полуавтоматического оборудования, а на объекте осуществляется только сращивание волокон.

Решения первой группы дают возможность предельно уменьшить стоимость покупных компонентов. В случае применения подхода второго типа несколько снижается общее время выполнения монтажных работ, резко уменьшается количество отходов и практически стопроцентно гарантируются стабильность параметров и высокое качество.

Основным отличительным признаком второй разновидности методов оконцевания является применение так называемого монтажного шнура или пигтейла. Монтажный шнур присоединяется к световоду ОК с помощью сварки или механического сплайса. Технология с применением монтажного шнура относится к группе решений с выполнением части технологических операций по оконцеванию еще на этапе изготовления. Поэтому применение в случае технологии сварки прецизионного оборудования с высокой степенью автоматизации вследствие использования управляющего контроллера позволяет получить рекордно низкое на сегодняшний день время оконцевания ОВ в сочетании с минимальными вносимыми потерями.

3.1. Клеевые и механические методы фиксации волокна в наконечнике

Процедуры оконцевания световодов ОК элементами ОР, при которых ОВ фиксируется в центрирующем наконечнике или замещающем его элементе, относятся в настоящее время к классическим. Применяемые на практике разновидности реализующих их технологий могут быть в общем случае разделены на две основные подгруппы: с клеевой и механической фиксацией.

Клеевые методы оконцевания ОВ занимали доминирующее положение среди других технологий на первых этапах развития техники СКС. Сейчас они являются стандартным методом установки ОР практически во всех СКС как европейских, так и американских производителей (см. табл. 4). Этому способствуют простота реализации, универсальность и минимальная стоимость, как применяемого инструмента, так и расходных материалов.

Компания производитель	Тип СКС (оптической)	Сплайсы подсистемы	Сварочные аппараты	Клеевой состав
AMP, США	NetConnect Core-link (Solarum)	—	Эпоксидная смола	AMP, США
Corning, США	LANscape CamSplice	Serie X7	—	Corning, США

Commscope, США	Systimax CSL (Optispeed)	—	Эпоксидная смола и двухкомпонентный клей	Commscope, США
АйТи, Россия	АйТи-СКС	Fujkura FSM-17 FSM-50	Двухкомпонентный клей	АйТи, Россия
Brand-Rex, Англия	Millennium (Helios)	Fiber Opticx	—	Brand-Rex, Англия
Nexans, Франция	Nexans Cabling System	—	Клей Hot Melt	Nexans, Франция
RiT Technologies, Израиль	Smart	—	Эпоксидная смола	RiT Technologies, Израиль
Panduit, США	PanNet	—	Двухкомпонентный клей	Panduit, США

Таблица 4. Технические средства для установки разъемов некоторых производителей СКС

3.2. Иммерсионные разъемы

Появление ОР иммерсионного типа, иногда называемых также механическими коннекторами, было вызвано стремлением к поиску альтернативы клеевой технологии и устранению ее недостатков при работе в первую очередь в полевых условиях. Наиболее серьезными из них являются достаточно сложная многоэтапная процедура реализации технологии, необходимость применения значительного количества расходных материалов, а также наличие в составе клея весьма агрессивных химических веществ.

Иммерсионные разъемы представляют собой комбинацию механического сплайса и предустановленного ОР, т.е. в соответствии с предложенной выше классификацией они относятся к решениям второй разновидности с частичным выполнением некоторых технологических операций в заводских условиях. Применение этих элементов позволяет радикально решить проблему уменьшения продолжительности монтажа при относительно невысокой стоимости оборудования и не слишком жестких требованиях в отношении квалификации монтажников. Такой эффект обеспечивается:

- устранением из процесса установки требующих повышенной аккуратности и занимающих много времени операций шлифовки и последующей полировки торцевой поверхности наконечника;
- применением в процессе монтажа комплекта хорошо отработанных технологических инструментов, в некоторых случаях снабжаемых элементами полуавтоматики.

3.3. Технологии механического и сварного сращивания

Далеко не все производители СКС предусматривают в составе оборудования оптических подсистем средства для сращивания ОВ. Обоснованием такого подхода служит сравнительно небольшая протяженность кабельных линий, что позволяет в подавляющем большинстве случаев создавать их с использованием одной строительной длины. Тем не менее, если производитель СКС считает необходимым иметь в составе своей системы средства для сращивания световодов, то он может взять на вооружение технологии механических сплайсов и сварки. Одним из главных преимуществ механических сплайсов (см. табл. 5) являются заметно меньшие затраты на технологическое оборудование для их установки.

Метод	Время установки, мин	Стоимость комплекта инструментов, долл. США	Стоимость конструкции, долл. США	Примечание
Клеевая технология	5–20	450–15000	2–13	–
Иммерсионные разъемы	2–3	700–1200	8–15	–
Механические сплайсы	1–3	1000	7–20	Необходим организатор сплайсов
Технология сварки	1–2	8000–30000	10–20	Необходим организатор защитных гильз

Таблица 5. Сравнительная характеристика различных методов оконцевания световодов

Благодаря этому применение механических сплайсов оказывается экономически выгодным даже при выполнении сравнительно небольшого количества сращиваний и оконцовок.

Технология сварки является в настоящее время самым распространенным среди строителей сетей связи общего пользования способом сращивания ОВ. Реализующие эту технологию сварочные аппараты используются:

- для соединения ОВ отдельных строительных длин кабелей в процессе строительства линий связи большой протяженности, а также во время монтажа разветвительных и переходных муфт;
- при оконцевании ОВ вилками ОР в случае применения монтажных шнуров.

По упоминавшимся выше причинам экономического и технического характера сварочные аппараты не получили повсеместного распространения в области структурированной проводки и включаются в состав штатного технологического оборудования лишь ограниченным количеством производителей СКС.

Заключение

В оптических подсистемах СКС используются разъемы исключительно контактного типа, построенные по схеме физического контакта и унифицированные с аналогичными изделиями для сетей связи общего пользования. Применение подобных компонентов позволяет снизить их стоимость и упрощает процесс производства, построения и эксплуатации структурированной кабельной проводки.

Наличие широкой гаммы серийных разработок разъемов из группы SFF (small form factor) позволяет достигнуть в оптической подсистеме плотности портов, по меньшей мере, не уступающей аналогичному параметру медной подсистемы. Дополнительным преимуществом элементной базы этой разновидности является то, что в нее еще на стадии разработки заложен принцип практически полной совместимости с разъемами модульного типа подсистемы на основе витой пары по посадочным местам и идентичности по эксплуатационным параметрам.

Основная масса типов серийных оптических разъемов реализует симметричную схему построения, что выгодно в первую очередь с эксплуатационной точки зрения, так как унифицирует конструкцию и технологию монтажа на линейных и шнуровых ка-

бельных изделиях. Определенное значение имеют также несколько более широкие функциональные возможности симметричных разъемов.

Использование основных компонентов оптического разъема оказывается технически выгодным в процессе конструирования разнообразных адаптеров. Изделия данной разновидности в массовом масштабе применяются при решении широкого круга задач, часто возникающих на этапах построения и эксплуатации структурированной проводки (изменение типа интерфейса, внесение в тракт передачи дополнительного затухания и т.д.).

В технике СКС находит применение значительное количество оригинальных технологий монтажа элементов разъемов, ориентированных на применение непосредственно на объекте установки кабельной системы. Это существенно упрощает и ускоряет процесс реализации оптической подсистемы. Каждый из внедренных в широкую инженерную практику методов позволяет гарантировать качество не только на уровне требований действующих редакций национальных и международных нормативно-технических документов, но и с определенным превышением пороговых значений.

Литература

1. Мальке Г., Гессинг П. Волоконно-оптические кабели. Основы, проектирование кабелей, проектирование систем. / Пер. с нем. 2-е изд., перераб. и доп. – Corning Cable Systems. – Новосибирск, 2001.
2. Волоконная оптика в локальных и корпоративных сетях связей. – М.: КомпьютерПресс, 1998.
3. Packaging guide: Фирменный материал компании Acome. – 2004.
4. ГОСТ 18690-82 (СТСЭВ 3227-81). Кабели, провода, шнуры и кабельная арматура. Маркировка, упаковка, транспортирование и хранение. Государственный комитет СССР по стандартам. – М.: Издательство стандартов, 1983.
5. Ларин Ю.Т. Программа разработки нормативно-технической базы оптических волокон и кабелей // Фотон-Экспресс. – 2003. – Сентябрь. – №4(30).
6. ГОСТ 5151-79. Барабаны деревянные для электрических кабелей и проводов. Технические условия. Государственный комитет СССР по стандартам. – М.: Издательство стандартов, 1986.
7. Семенов А.Б. Оптические разъемы // Фотон-Экспресс. – 2005. – №4(44).
8. Московченко А. Оптические коннекторы SFF – достойный ответ на современные и перспективные запросы телекоммуникационной индустрии // Электронные компоненты. – 2003. – № 1.

ИССЛЕДОВАНИЕ НАНОКОМПОЗИТНЫХ СТРУКТУР НА ОСНОВЕ КРЕМНИЯ И ФОСФАТНОГО СТЕКЛА НА КРЕМНИЕВЫХ ПОДЛОЖКАХ

И.И. Стройков

Научный руководитель – д.т.н., профессор А.М. Скворцов

В работе проведены исследования нанокompозитных пленок на основе кремния, диоксида кремния и фосфатного стекла. Разработана лабораторная технология получения нанокompозитных пленок. Пленки изготавливались на термически окисленных монокристаллических кремниевых пластинах. Исследованы структурные, а также электрофизические особенности и характеристики, полученных нанокompозитов.

Введение

Нанокompозиты (НК) – новые материалы, потенциально обладающие уникальным набором свойств, не встречающихся ни у одного природного материала. Принципиальное отличие НК от других композитов состоит в высокой однородности свойств, обеспеченной аморфностью матрицы и наноразмерами формируемых в ней частиц.

Научный интерес к НК структурам и материалам связан, прежде всего, с ожиданием различных свойств размерных эффектов наночастиц или наноструктур, размеры которых соизмеримы или меньше, чем характерный корреляционный масштаб того или иного физического явления или характерная длина, фигурирующие в теоретическом описании какого-либо свойства или процесса (например, длина свободного пробега электронов, дебройлевская длина волны, размер магнитного домена в ферромагнетиках и пр.).

Проблема получения тонкодисперсных порошков металлов, сплавов, соединений и сверхмелкозернистых материалов из них, предназначенных для различных областей техники, обсуждается давно. В последнее десятилетие интерес к этой теме существенно возрос, так как обнаружилось, что уменьшение размера частиц материала ниже некоторой пороговой величины может проводить к значительному изменению свойств. Такие эффекты появляются, когда средний размер кристаллических зерен не превышает 100 нм, и наиболее отчетливо наблюдается, когда размер зерен менее 10 нм.

Управление фундаментальными свойствами твердых тел (полупроводники, металлы, полимеры и т.д.), основанное на синтезировании в их объеме наноразмерных фаз выделений, кристаллитов, дефектных структур или формировании на поверхности пленочных наноструктур, в настоящее время составляет одну из главных проблем ведущих научных центров мира, работающих в направлении нанотехнологий [1].

Наночастицы и нанослои широко применяются в производстве современных микроэлектронных устройств. Нанокристаллические материалы представляют собой особое состояние конденсированного вещества – макроскопические ансамбли ультрамалых частиц с размерами до нескольких нанометров. Необычные свойства этих материалов обусловлены как особенностями отдельных частиц (кристаллитов), так и их коллективным поведением, зависящим от характера взаимодействия между наночастицами [2].

Методика эксперимента

Технологию получения нанокompозитных пленок на кремниевых подложках стоит условно разделить на несколько основных этапов. Она включает в себя:

- приготовление порошков используемых материалов;
- приготовление суспензии на основе полученных порошков;
- нанесение суспензии на подложку;
- формирование пленки НК путем спекания.

Для приготовления порошка кремния использовались кремниевые не окисленные пластины, предварительно прошедшие несколько этапов очистки:

- промывка с использованием чистящих средств (порошок, мыло и т.д.);
- травление в кипящем растворе калиевой щелочи KOH;
- промывка в дистиллированной воде;
- сушка в термостате при температуре $T = 100^{\circ}\text{C}$.

Получение порошка кремния производилось с помощью метода механосинтеза. Простейший аппарат для измельчения – шаровая вращающаяся мельница – представлен на рис. 1. Она представляет собой металлический цилиндрический барабан, внутри которого находятся размольные тела – стальные шары (при изготовлении порошка использовались два больших шара $d = 18$ мм и три маленьких $d = 13$ мм) [3].

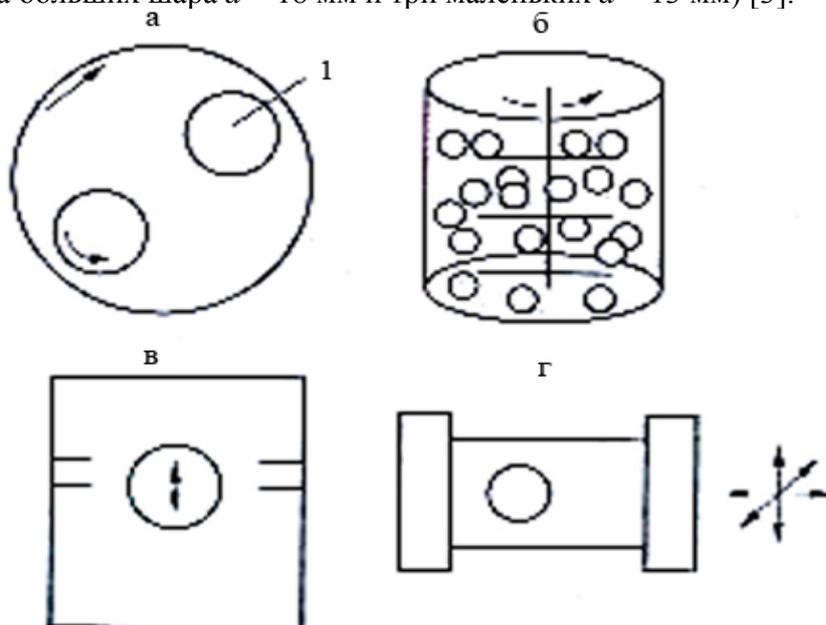


Рис. 1. Типы машин для шарового размоль (а – планетарного типа; б – аттрактор; в – одномерная вибрационная машина; г – трехмерная вибрационная машина) [4]

Измельчение кремния производилось на шаровой мельнице планетарного типа (рис. 1, а). Процесс измельчения проводился в течение 180 часов. Определение размера частиц было выполнено с помощью седиментационного метода (метод осаждения). Было определено, что размер частиц варьируется от сотен нанометров до микронных размеров. Для приготовления порошка стекла использовалось фосфатное стекло, легированное эрбием. Изготовление порошка стекла проходило по схожей схеме. Крупный кусок стекла был изначально измельчен и загружен в мельницу.

Следующим этапом технологии является приготовление водной суспензии. В качестве основы суспензии могут быть использованы различные щелочи, кислоты, масла и т.д. В нашем случае основой суспензии выбрана дистиллированная вода.

Были проведены эксперименты по окислению кремниевого порошка при комнатной температуре. На измеренную массу порошка кремния (насыпанного на пластину)

была нанесена вода. В течение некоторого времени мокрый порошок выдерживался, а потом была произведена просушка. Полученный образец с высушенным порошком кремния был взвешен, и его масса оказалась больше, чем масса первоначально сухого порошка. Таким образом, можно сделать вывод, что кремний взаимодействует с водой при комнатной температуре и окисляется, а порошинки объединяются между собой этим окислом.

В данной работе в качестве порошковой смеси использовались порошки кремния и фосфатного стекла в весовых соотношениях 1:1. Смешивание порошков производилось в фарфоровой ступке – истиранием с помощью фарфорового песта. Затем была изготовлена водяная суспензия. Учитывая то, что кремний, в отличие от стекла, плохо смачивается водой, в суспензию было добавлено поверхностно активное вещество. Далее осуществлялось отстаивание суспензии в течение 20 мин, чтобы опустились на дно наиболее крупные частицы кремния и стекла.

Из полученного столба водяной взвеси высотой 25 мм с помощью шприца бралась суспензия с высоты 10 мм и наносилась на поверхность подложки. В качестве подложки в работе использовались различные кремниевые пластины (пластины с разными типами проводимости и толщинами окисла), предварительно прошедшие этапы очистки. Высота капли суспензии на подложке составляла около 2 мм. Далее следовала просушка полученного образца, для удаления влаги с поверхности, в термостате при температуре $T = 100^{\circ}\text{C}$ в течение 30–40 мин. Диаметр полученных капель составлял 2–3 мм.

На рис. 2 представлены микрофотографии полученных образцов.

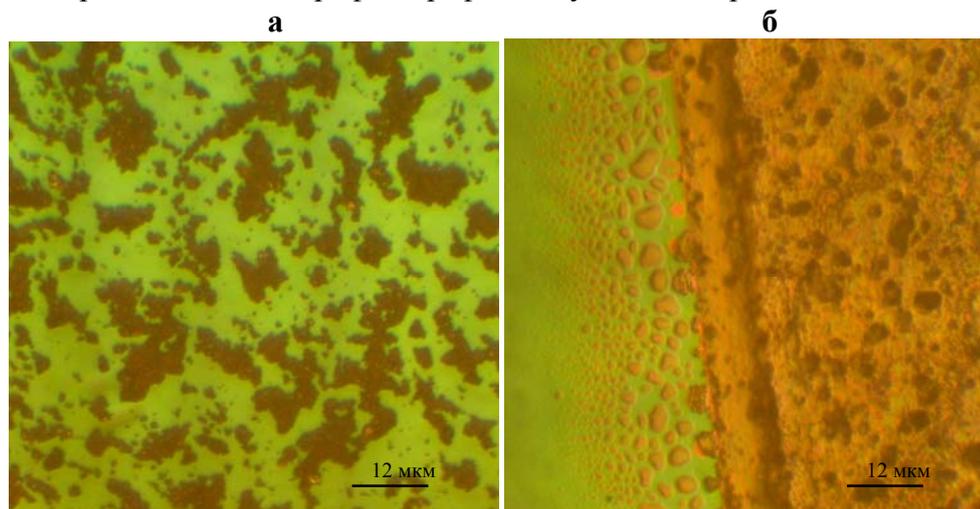


Рис. 2. Микрофотографии осажденной суспензии на кремниевую подложку при 800-кратном увеличении: а – центр пятна, б – на краю пятна

На микрофотографиях с изображением полученного высушенного пятна видно, что произошло неравномерное осаждение порошков, и концентрация порошка на краю существенно отличается от центральной области. Причиной такого распределения частиц является неравномерный процесс высыхания выпуклой капли водяной суспензии.

Следующим этапом формирования НК-пленки является термоотжиг. Были опробованы различные концентрации порошков. В ходе экспериментов выяснилось, что образцы, в которых преобладает порошок кремния – не спекаются. Этот эффект можно объяснить тем, что кремний имеет значительно большую температуру плавления, чем стекло. В итоге суспензия из порошка кремния и фосфатного стекла была изготовлена в пропорциях 1:1. Спекание производилось в диффузионных печах СДО-3 в течение 4 часов при температуре в пределах $1000\text{--}1100^{\circ}\text{C}$. Разогрев и остывание подложек происходило вместе с печью [3].

Результаты исследований

Структурные характеристики

Пленки после процессов спекания и лазерного облучения представляют собой композитные структуры. На рис. 3 представлены микрофотографии структур, полученных при спекании. На рис. 3, а, видно, что в процессе спекания под воздействием высокой температуры частицы стекла расплавлены. Темные пятна представляют собой частицы кремния. Так как образец спекался в течение достаточно продолжительного времени (4 часа непосредственного спекания при температуре 1000°C и время полного остывания печи – 2,5 часа), то после процессов плавления и последующего остывания произошло высыхивание, и полученная стеклянная пленка растрескалась.

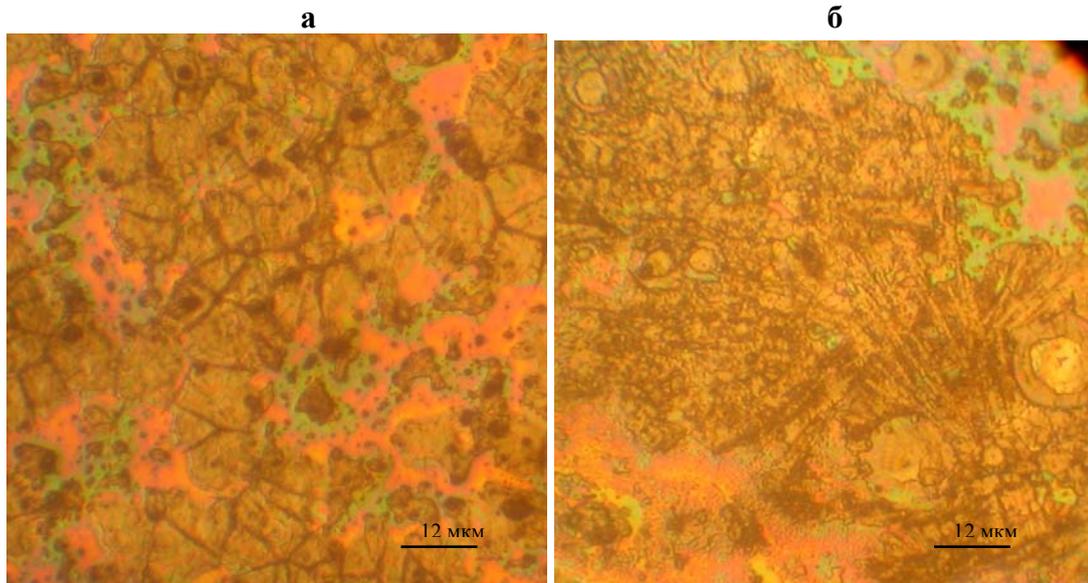


Рис. 3. Микрофотографии структур, полученных спеканием при 800-кратном увеличении: а – область вблизи края пленки НК, б – край пленки НК

Процесс плавления частиц стекла влечет за собой перемещение материала, что видно на рис. 3, б. Сравнивая вид образцов до и после высокотемпературного отжига, можно заметить, что произошло перераспределение концентраций частиц и образовались разреженные области. Таким образом, следует предположить, что рельефы на рис. 3, б, образовались вследствие перемещения частиц кремния вместе с расплавленным стеклом.

В процессе проведения исследований был произведен эксперимент с отжигом нанесенной высушенной суспензии с помощью лазерного облучения. Пленки суспензии были облучены интерференционным непрерывным пучком Nd:YAG лазера диаметром 50 мкм с длиной волны 1,06 мкм, с мощностью, достаточной для расплавления частиц, с шагом 50 мкм. Процесс облучения производился при комнатной температуре в воздухе. Полученные лазерным отжигом пленки имели более равномерную структуру, в которой при определенных условиях могли быть сформированы отчетливо видимые шарики кремния. Этот предварительный эксперимент указывает на целесообразность применения лазера для окончательного формирования НК-пленки.

Вольт-фарадные характеристики

По мере снятия вольт-фарадных характеристик было установлено, что зависимости, полученные в центре пятна и на самой подложке, совпадают. Это объясняется тем, что в центре полученной области концентрация частиц мала, и они находятся на большом расстоянии друг от друга.

На рис. 4 и рис. 5 представлены вольт-фарадные характеристики, полученные на краю пятна, нанесенного на кремниевую подложку р- и n- типа, соответственно. НК- пленки сформированные на р – подложке (рис. 4) характеризуются резкой зарядкой емкости области пространственного заряда и диэлектрика в области обогащения.

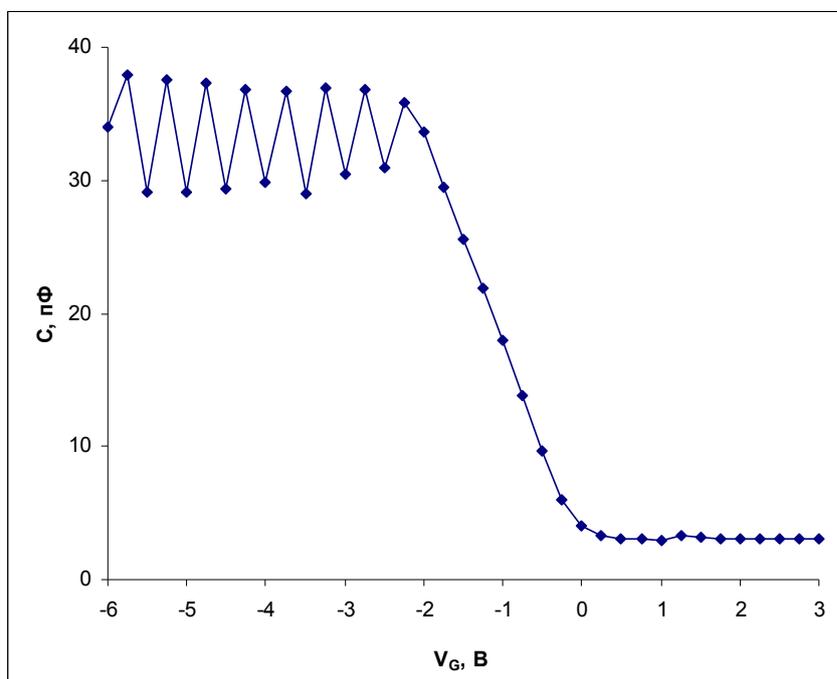


Рис. 4. Вольт-фарадная характеристика НК-пленки на кремниевой подложке р-типа при частоте сигнала 1 МГц

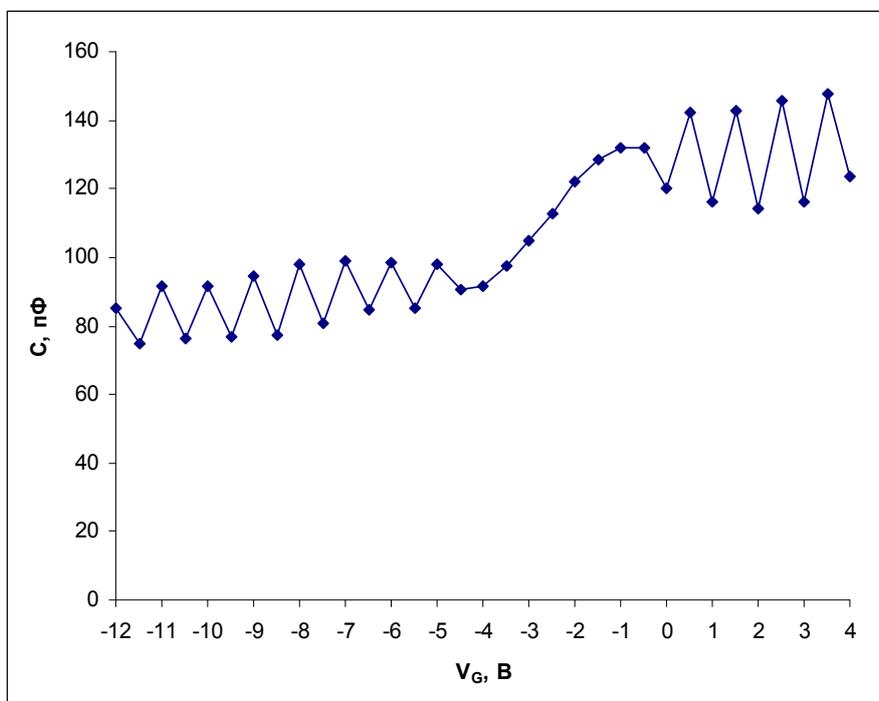


Рис. 5. Вольт-фарадная характеристики НК-пленки на кремниевой подложке n-типа при частоте сигнала 1 МГц

Таким образом, как следует из вольт-фарадных характеристик, при формировании на поверхности SiO_2 пленки нанокompозита на основе фосфатных стекол и кремния приводит к изменению стандартных кривых ВФХ, например, на подложках р-типа проводимости в области обогащения. На ВФХ структур с пленками НК на типичную кривую изменения емкости накладываются дополнительные более высокочастотные изменения емкости ΔC , которые свидетельствуют о перезарядке некоторых центров захвата носителей. Этими центрами, по-видимому, являются нанокристаллы кремния, встроенные в пленку нанокompозита.

На вольт-фарадных характеристиках подложек п-типа эффект перезарядки емкости на включениях нанокристаллов кремния наблюдается на обоих участках зависимости (рис. 5). Для выяснения причин этого эффекта требуются дополнительные исследования.

Вольт-амперные характеристики

В данной работе снимались вольт-амперные характеристики (ВАХ) для исследования электрической прочности (пробивных напряжений) полученных НК-пленок и исследования процессов, происходящих после пробоя.

При снятии ВАХ для обеспечения омического контакта зондов и точек на НК-пленке использовался сплав InGa . Характеристики снимались следующим образом. Предварительно очищенная от окисла обратная сторона пластины (свободная от образцов) смачивалась InGa и устанавливалась на медную площадку, к которой подводился первый контакт. Для получения второго контакта на участок с НК-пленкой в нужную точку наносилась капля InGa , и в эту каплю помещался тонкий медный зонд, не соприкасающийся непосредственно с самой структурой. Сначала подавалось напряжение, при котором происходил пробой нанокompозитной пленки и диэлектрика. Затем снималась послепробойная вольт – амперная характеристика (ПВАХ).

На рис. 6 представлены типичные ВАХ, снятые на окисле подложки в стороне от нанокompозитной пленки.

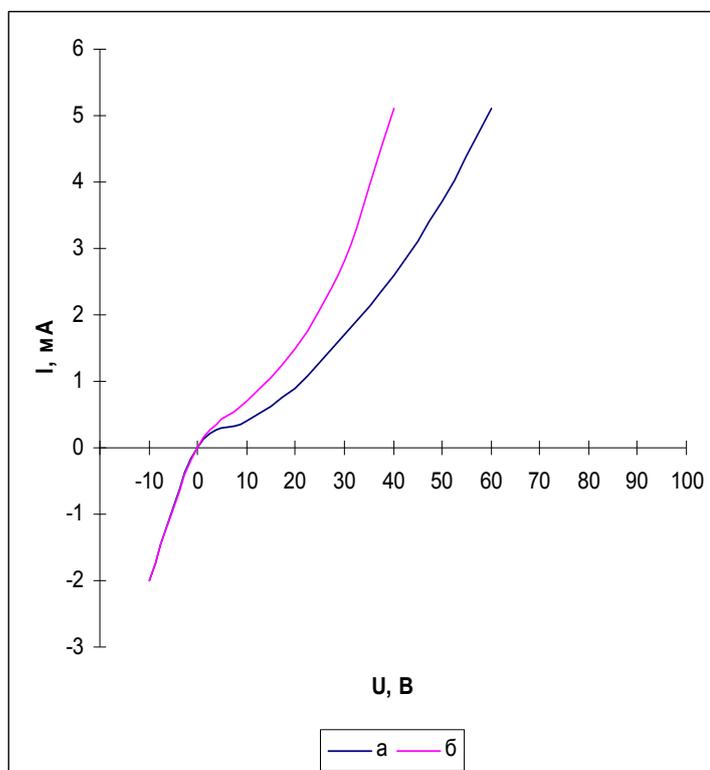


Рис. 6. Вольт-амперные характеристики окисленного кремния р-типа, снятые после пробоя диэлектрика; а, б – различные точки на окисле

На рис. 7 изображены наиболее типичные характеристики, полученные при установлении зонда на край НК пленки, т.е. в область с высокой концентрацией спеченного порошка.

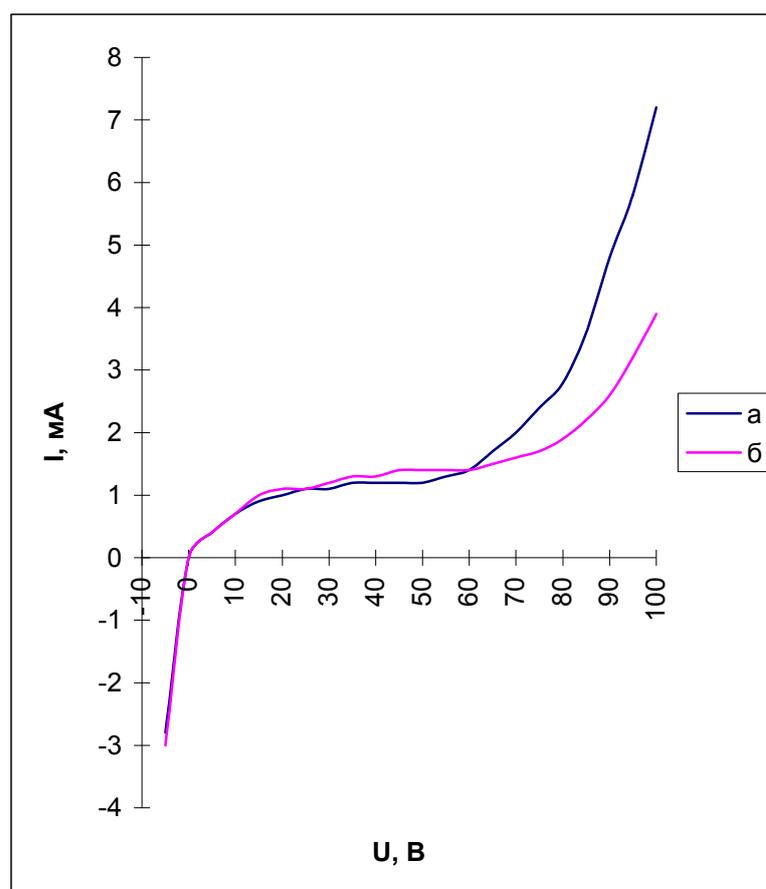


Рис. 7. Вольт-амперные характеристики полученных НК-пленок, снятые после пробоя пленки и диэлектрика; а, б – различные точки на НК-пленке

Наибольший интерес вызывает область с очень медленным возрастанием тока (20–70 В). Таким образом, наблюдаемая на этой части кривой проводимость, по-видимому, объясняется наличием в пленке НК окисленных частиц кремния. Увеличение проводимости может быть связано с локальными пробоями окисла между кластерами кремния. При достижении некоторых критических значений напряжения (70–80 В) кривая принимает вид, характерный для пробоя диэлектрика.

В дальнейшем планируется провести более детальное исследование полученных НК-пленок, а также изучение нанокompозитных структур, полученных с помощью лазерного облучения.

Заключение

В данной работе были получены и исследованы нанокompозитные пленки на монокристаллических подложках кремния. В результате проделанных экспериментов и опытов получены следующие результаты.

- Показаны структурные особенности.
- Исследованы вольт-амперные и вольт-фарадные характеристики.
- Показано, что наличие НК-пленок на поверхности SiO_2 вносят в стандартные ВАХ и ВФХ изменения, которые, по-видимому, связаны с перезарядкой нанокристаллов кремния.

- Показана целесообразность использования лазерного облучения для окончательного формирования НК-пленки на основе микропорошков.

Следует отметить, что в дальнейшем планируется более детальное и широкое исследование нанокompозитных пленок, полученных различными методами, в которое будут входить спектральные и фотолюминесцентные характеристики.

Литература

1. Гусев А.И., Ремпель А.А. Нанокристаллические материалы. – М.: ФИЗМАТЛИТ, 2001. – 224 с.
2. Мальцев П.П. Наноматериалы. Нанотехнологии. Наносистемная техника. – М.: ТЕХНОСФЕРА, 2006. – 152 с.
3. Стройков И.И. Разработка лабораторной технологии получения нанокompозитных пленок на кремниевых подложках // Научно-технический вестник СПбГУ ИТМО. – 2007. – Вып. 40. – С 90–94.
4. Валиев Р.З., Грабовецкая Г.П., Колобов Ю.Р. Зернограничная диффузия и свойства наноструктурных материалов. – Новосибирск: Наука, 2001. – 232 с.

АНАЛИЗ СОВРЕМЕННОЙ ПАТЕНТНОЙ ЛИТЕРАТУРЫ ПО МИКРОМЕХАНИЧЕСКИМ АКСЕЛЕРОМЕТРАМ

А.Ю. Буданова, В.А. Крылов

Научный руководитель – д.т.н., профессор В.Л. Ткалич

В статье проведен анализ отечественных и зарубежных работ по микромеханическим акселерометрам, который позволил определить основные направления исследований по повышению точности работы данных устройств. По итогам патентно-информационного поиска выявлена высокая актуальность дальнейшего развития микромеханических чувствительных элементов, а также определена перспективность теоретического исследования акселерометров еще на стадии проектирования.

Введение

Микромеханический акселерометр (ММА) – это чувствительный элемент, широко используемый в различных системах ориентации, стабилизации, наведения и навигации. Несмотря на то, что основным назначением ММА является измерение ускорения, эти датчики способны также измерять наклон, движение, положение, силу ударов и вибрацию. ММА используются в производстве компьютерной, строительной, геодезической и медицинской продукции. Изделия на базе микротехнологий становятся предметами повседневной жизни. Их все больше применяют в средствах обеспечения жизнедеятельности человека – от систем безопасности и биоконтроля до аппаратуры автоматизации дома, промышленного контроля, мониторинга окружающей среды и, конечно, бытовой техники и развлекательных систем [1].

В основу работы чувствительного элемента (ЧЭ) инерционных ММА могут быть положены различные физические эффекты: емкостной, пьезорезистивный, пьезо- и сегнетоэлектрический, электромагнитный, оптический, туннельный.

В простейших инерционных ММА применяют ЧЭ, реализующие принцип емкостного преобразования. Преимущества емкостного метода измерений состоят в следующих факторах:

- высокая линейность;
- высокая чувствительность;
- легко осуществимая самокалибровка и самотестирование;
- легкая реализация цепей обратной связи;
- совместимость с КМОП-технологией;
- широкий диапазон рабочих температур.

По принципу действия ММА можно разделить на приборы прямого преобразования (ПП) и компенсационного (КП) (табл. 1). Чувствительный элемент ММА ПП непосредственно передает информацию о входной величине на корректирующее звено, и при этом все погрешности измерительного тракта присутствуют в выходном сигнале датчика (рис. 2). ММА КП частично или полностью уравнивают измеряемую величину с помощью отрицательной обратной связи (ООС), т.е. реализуют силовую разгрузку ЧЭ с помощью выходного сигнала, подаваемого на элемент компенсации. В последнем случае точность прибора зависит в основном от элемента компенсации – датчика силы или момента обратной связи. В приборах прямого измерения, напротив, точность зависит от всех узлов, что приводит к усложнению схемотехники таких конструкций [2, 3].

Нормируемыми характеристиками ММА, как правило, являются резонансная частота, диапазон входного преобразуемого сигнала, уровень шума, осевая чувствительность, дрейф, линейность, динамический диапазон характеристики преобразования, устойчивость к ударным воздействиям, мощность потребления, температурный диапазон. Целью данной статьи является проведение патентного поиска и анализа по основным

усовершенствованиям ММА за период 2000–2007 гг. и выявление основных направлений исследований по модернизации этого устройства.

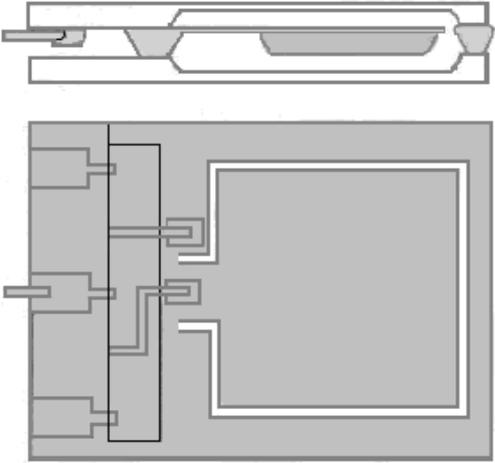
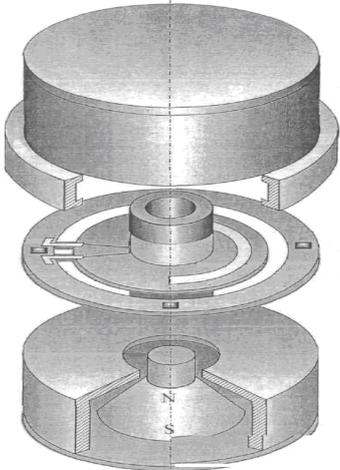
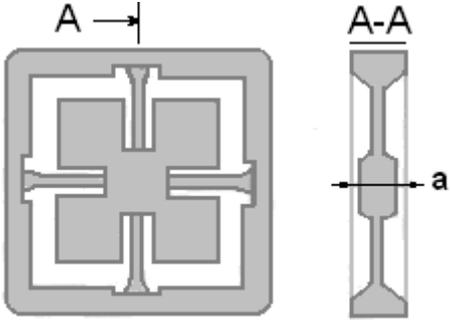
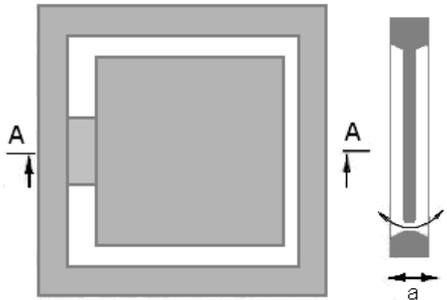
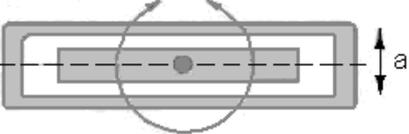
По принципу действия	
прямого преобразования	компенсационного преобразования
 <p>Конструкция балочного акселерометра без обратной связи</p>	 <p>Конструкция гибридного акселерометра с магнитоэлектрической обратной связью</p>
тип обратной связи	
	электростатическая
	магнитоэлектрическая
По типу чувствительного элемента	
осевой ММА	маятниковый ММА
 <p>Принципиальная схема ЧЭ</p>	 <p>Принципиальная схема ЧЭ</p>
вид измеряемого движения	
линейные ММА	угловые ММА
 <p>Принципиальная схема ЧЭ ММА с двумя режимами работы</p>	
тип основания	
на колеблющемся основании	на поступательно перемещающемся основании

Таблица 1. Классификация микромеханических акселерометров

1. Изменение конструкции рамки чувствительного элемента ММА

Рассмотрим микромеханический акселерометр с чувствительным элементом маятникового типа (см. таблицу 1). Известно, что чувствительность конструкции к контактным напряжениям определяется площадью контакта и удаленностью места заделки упругого подвеса от силового контакта:

$$\frac{\sigma_{dj}}{P} = \frac{1-\nu^2}{y_0} \int_s \frac{dS_j}{r}, \quad (1)$$

где ν – коэффициент Пуассона; P – давление на контакт; r – расстояние от точки приложения сосредоточенной силы до заданного сечения; S – площадь контакта упоров; y_0 – толщина упора; σ_{dj} – напряжение в j -м месте стыка упругих подвесов и упоров. Недостатком известных устройств [4, 5] является нестабильность смещения нуля вследствие высокого уровня контактных напряжений, возникающих в местах крепления упоров, следовательно, снижается точность прибора в целом. Для повышения точности прибора необходимо уменьшить нестабильность нуля за счет снижения влияния контактных напряжений.

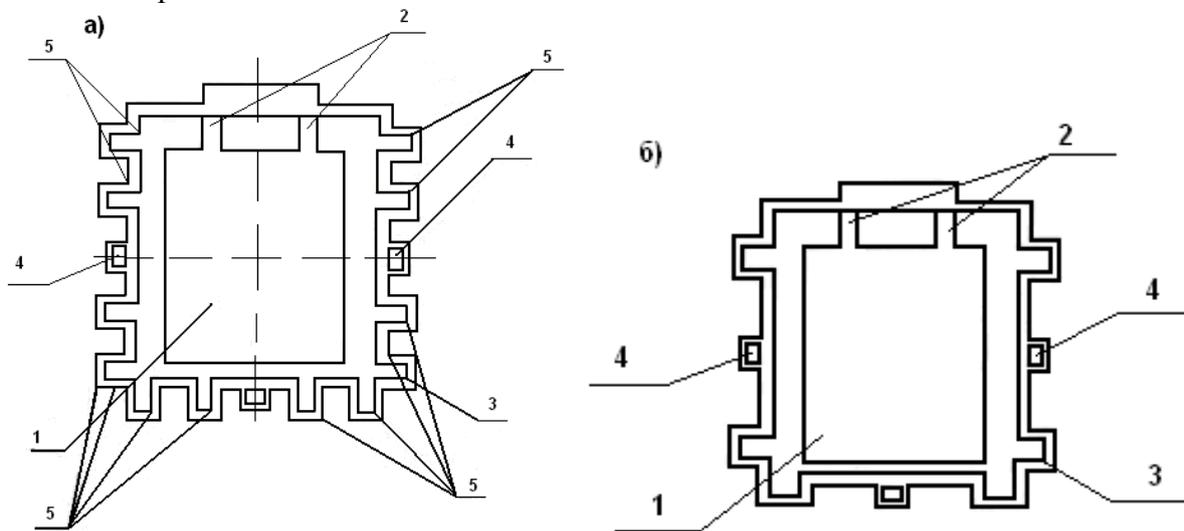


Рис. 1. Принципиальная схема ЧЭ ММА: а – патент РФ № 2284528; б – патент РФ № 2265856. 1 – ЧЭ маятникового типа; 2 – упругие подвесы; 3 – рамка ЧЭ; 4 – упоры; 5 – изгибы рамки

Согласно изобретению [6], рамка чувствительного элемента на участках между упорами, а также на участках между упорами и упругим подвесом выполнена в форме меандра с количеством изгибов n , при этом внешний и внутренний изгибы меандра одинаковы и симметричны относительно продольной оси симметрии чувствительного элемента, а ширина рамки постоянна (рис. 1, а). Меандр строго симметричен относительно продольной оси симметрии чувствительного элемента, тем самым увеличивается расстояние от мест крепления упоров до упругого подвеса, а габаритные размеры чувствительного элемента остаются неизменными, и, следовательно, исходя из выражения (1), уменьшается влияние контактных напряжений на упругий подвес. Количество изгибов определяется разрешающей способностью литографии и анизотропного травления, а также ориентацией исходной пластины. На рис. 1, б представлен вариант с 4 изгибами [7]. Кроме того, рамка равномерно деформируется под воздействием внешних факторов, а симметричное расположение компенсирует до минимума вредные воздействия.

2. Использование систем терморегулирования в ММА

Практически все типы инерциальных датчиков чувствительны к изменениям температуры окружающей среды, что приводит к температурному дрейфу нуля и изменению чувствительности. Этот недостаток присущ и ММА. Эффективным методом уменьшения влияния температуры окружающей среды является введение систем стабилизации температуры (ССТ) микромеханического ЧЭ.

Для уменьшения температурной погрешности акселерометров применяют методы термокомпенсации или термостатирования.

Метод термокомпенсации заключается в том, что сигнал акселерометра корректируется сигналом от датчика температуры по какому-либо закону. Термокомпенсация может быть программной или аппаратной.

- Специальные нагревательные элементы, способные рассеивать мощность в датчиках. Например, в ММА [8] введен резистивный нагревательный элемент, который позволяет за счет ручной регулировки уменьшить градиенты температуры в этом ЧЭ в начальные моменты времени после включения, однако не обеспечивает стабилизации температуры ММА. Соответственно, недостатком устройства [8] является снижение точности чувствительного элемента при изменениях температуры окружающей среды. В процессе работы дополнительно рассеивается мощность на регулировочном резисторе, поскольку резистивный нагреватель является нелинейным элементом. При линейном изменении тока через резистивный элемент рассеиваемая на нем мощность изменяется по квадратичному закону. Наличие в замкнутом контуре системы регулирования такого нелинейного элемента существенно снижает как статические, так и динамические характеристики системы. Также недостатком является использование пассивной ССТ (ручная настройка, выполняемая один раз).
- В изобретении [9] используется встроенная система стабилизации с корректирующим звеном для автоматического управления нагревательным элементом. На рис. 2 приведена блок-схема устройства стабилизации температуры ЧЭ. Датчик температуры 3 преобразует температуру чувствительного элемента 1 в электрический сигнал, который корректирующим звеном 4 усиливается и преобразуется в сигнал управления нагревательным элементом 2. Нагревательный элемент 2 изменяет рассеиваемую в чувствительном элементе 1 мощность в зависимости от сигнала датчика температуры 4 и поддерживает температуру чувствительного элемента 1 постоянной.

В рассмотренном патенте [9] при использовании в качестве нагревателей транзисторов типа КТ814 было получено уменьшение влияния температуры окружающей среды в 30 раз. Это обусловлено тем, что при симметричном размещении нагревательных элементов по периферии датчика температуры достигается стабилизация средней по объему чувствительного элемента температуры. Подобное расположение нагревательных элементов и датчика температуры обеспечивает стабилизацию средней температуры всего кристалла кремния, в котором реализован ЧЭ датчика. Теплоизоляция ЧЭ реализуется при помощи теплоизолирующего основания и использования теплоизолирующей крышки, что позволяет снизить потребляемую при стабилизации температуры мощность и тем самым повысить экономичность устройства.

При использовании **метода термостатирования** акселерометр нагревается до заданной температуры статирования, которая в течение работы поддерживается постоянной. Для уменьшения температурной погрешности ММА ADXL105 используется метод термостатирования. Разработчиками Analog Devices Inc [10] предлагается ССТ, представленная на рис. 3.

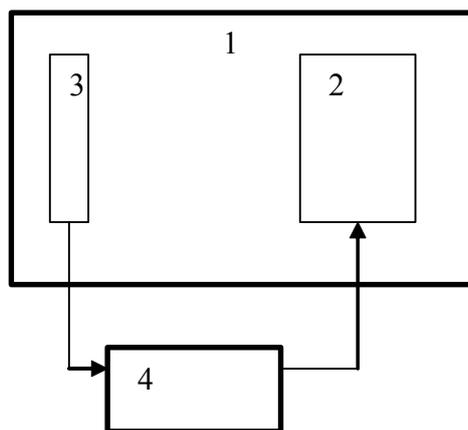


Рис. 2. Блок-схема устройства стабилизации температуры

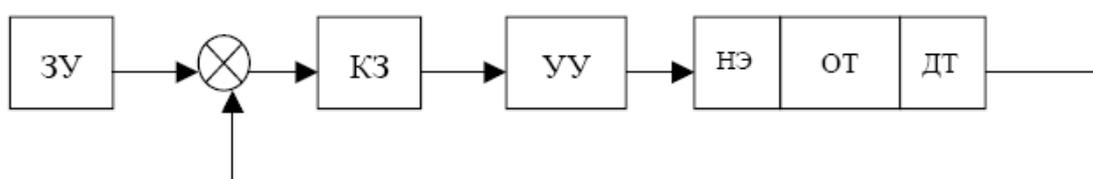


Рис. 3. Функциональная схема системы стабилизации температуры

	Термокомпенсация		Термостатирование
	Аппаратная	Программная	
Недостатки	<ul style="list-style-type: none"> – Нелинейная зависимость выходного напряжения акселерометра от напряжения с датчика температуры – Гистерезис 	<ul style="list-style-type: none"> – Высокая стоимость – Увеличение габаритов 	<ul style="list-style-type: none"> – Увеличение потребляемой мощности – Увеличение времени готовности акселерометра
Достоинства	<ul style="list-style-type: none"> – Готовность датчика к работе сразу после включения – Низкое потребление тока включения 		<ul style="list-style-type: none"> – Исключается гистерезис и нелинейность датчика

Таблица 2. Достоинства и недостатки систем регулирования температуры в микромеханических акселерометрах

Система стабилизации температуры включает в себя датчик температуры (ДТ), задающее устройство (ЗУ), корректирующее звено (КЗ), нагревательный элемент (НЭ), усилительное устройство (УУ) и объект термостатирования (ОТ). При создании ССТ выбор типа и места установки ДТ и НЭ является, как правило, самой сложной задачей. Датчик температуры в акселерометре ADXL105 встроен, а для НЭ имеется только одна верхняя поверхность. Использование термостатирования позволило уменьшить дрейф нуля акселерометра. Корректировка сигнала акселерометра при изменении температуры на 35°C позволила уменьшить температурный дрейф термостабилизированного ММА в 2 раза. Вышеперечисленные методы уменьшения температурной погрешности имеют свои достоинства и недостатки, которые отражены в табл. 2.

3. Изменение конструкции чувствительного элемента ММА

Рассмотрим ММА с подвесом маятника, разработанного в Draper Laboratory (США). Маятник 1 (рис. 4) представляет собой прямоугольную в плане пластину постоянной толщины, подвешенную на паре торсионных упругих элементов 2. Маятниковый эффект достигается асимметрией подвеса пластины на торсионах. Под пластиной размещена пара электродов 4, изменение емкости которых происходит при вращении пластины на торсионах при действии ускорения. Под пластиной размещены также электроды 3 контура компенсации [2, 11]. Также в [12] представлена конструкция ММА, содержащая маятниковый ЧЭ, имеющий два плеча.

В представленных конструкциях ось симметрии фигуры инерционной массы не совпадает с осью, проходящей через торсионы подвеса, что не обеспечивает симметричность элементов емкостной системы съема перемещений и элементов электростатического датчика момента по отношению к его поверхности инерционной массы, обращенной к электронной плате. При этом не симметрируются так называемые «паразитные» электрические емкости. Это оказывает влияние на изменение уровня нулевого сигнала емкостной системы съема перемещений, а также на изменение ее масштабного коэффициента.

В ММА [12] маятниковый подвес обеспечен путем смещения оси симметрии фигуры инерционной массы относительно оси подвеса, проходящей через его торсионы, при этом на внешней, по отношению к электрической плате, стороне большего плеча инерционной массы размещается дополнительный груз, увеличивающий общую массу, увеличивая тем самым изгибные деформации торсионов подвеса по «паразитным» степеням свободы при линейных перегрузках. В результате снижается устойчивость и прочность ММА к линейным перегрузкам и механическим ударам.

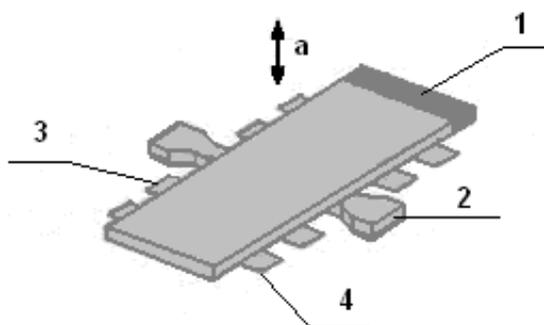


Рис. 4. Принципиальная схема маятникового ММА на торсионах

Для устранения вышеперечисленных недостатков разработчики ММА [13] выполнили ось симметрии фигуры инерционной массы, которая совпадает с осью, проходящей через торсионы подвеса. Маятниковый подвес обеспечивается удалением массы на внешней стороне одного плеча инерционной массы. На ней сделаны углубления, при этом для обеспечения высокой жесткости на изгиб и кручение указанного плеча его поверхность выполнена с ребрами жесткости (рис. 5). Совмещение оси симметрии фигуры инерционной массы с осью, проходящей через торсионы подвеса, обеспечивает симметричность элементов емкостной системы. При этом симметрируются так называемые «паразитные» электрические емкости, что позволяет снизить их влияние на изменение уровня нулевого сигнала емкостной системы съема перемещений, а также на изменение ее масштабного коэффициента.

Выполнение маятникового подвеса удалением части массы со стороны одного плеча инерционной массы снижает ее массу, уменьшая тем самым изгибные деформации торсионов подвеса по «паразитным» степеням свободы при линейных перегрузках.

В результате повышается устойчивость и прочность ММА к линейным перегрузкам и механическим ударам. Выполнение плеча инерционной массы с ребрами жесткости также снижает ее изгибные и крутильные деформации при перегрузках.

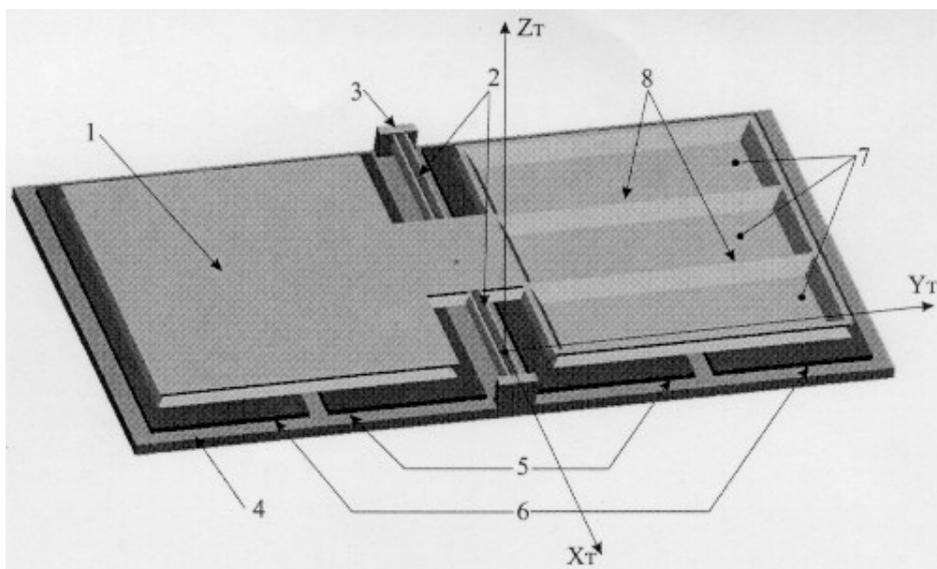


Рис. 5. Конструктивная схема ММА: 1 – инерционная масса; 2 – торсионы; 3 – корпус; 4 – электрическая плата; 5,6 – электроды; 7 – углубления; 8 – ребра жесткости

Заключение

Материалы российских журналов («Микросистемная техника», «Датчики и системы») и информационные материалы зарубежных сайтов [10, 14, 15] отражают новейшие разработки и способы производства интегральных датчиков для различных областей техники. В настоящее время разработчики стремятся повысить точность и чувствительность микромеханических акселерометров. Проведенный патентный поиск и последующий анализ усовершенствований ММА позволил выделить основные направления исследования данных устройств:

- изменение конструкции ММА;
- введение устройств стабилизации температуры ММА.

В результате анализа отечественных и зарубежных работ по микромеханическим акселерометрам и проведенного патентно-информационного поиска выявлена высокая актуальность дальнейшего развития интегральных датчиков на базе МЭМС-технологии.

Литература

1. Гольцева М. Юдинцев В. МЭМС – Здесь, там, везде. Большие рынки малых устройств // Электроника: Наука, Технология, Бизнес. – 2007. – № 1. – С.114–119.
2. Распопов В.Я. Микромеханические приборы: учебное пособие. – М.: Машиностроение, 2007. – 400 с.
3. Долгов А.Н. Электронные блоки для интегральных акселерометров прямого измерения и компенсационного типа: автореф. дис. ... канд. техн. наук: 05.11.03 / Долгов А.Н. – Нижний Новгород, 2006. –16 с.
4. Пат. № 2047863, РФ, Чувствительный элемент акселерометра. Баженов В.И. Вдовенко И.В. и др. Дата подачи заявки 1994.03.15. Дата публикации 1995.11.10.

5. Пат. № 2020484, РФ. Маятниковый компенсационный акселерометр с упругим подвесом. Горбачев Н.А., Рязанов В.А. Дата подачи заявки 1991.03.01. Дата публикации 1994.09.30.
6. Пат. № 2284528, РФ. Микромеханический датчик линейных ускорений. Миронов С.Г., Горинов С.А. Дата подачи заявки 2005.04.04. Дата публикации 2006.09.27.
7. Пат. № 2265856, РФ, G01P15/08. Микромеханический датчик линейных ускорений. Паршин В.А., Миронов С.Г., Горинов С.А. Дата подачи заявки 2004.07.06. Дата публикации 2005.12.10
8. Пат. № 2120639, РФ. Акселерометр. Курносков В.И., Андрюхин А.И. Дата подачи заявки 2002.06.18 Дата публикации 2005.10.20
9. Пат. № 2244936, РФ. Устройство стабилизации температуры микромеханического чувствительного элемента. Некрасов Я.А. Дата подачи заявки 1999.11.30. Дата публикации 2005.01.20
10. Сайт компании Analog Devices Inc. – Режим доступа: <http://www.analog.com>, свободный. – Загл. с экрана.
11. Сайт компании Draper Inc.– Режим доступа: <http://www.draper.com/mems>, свободный. – Загл. с экрана.
12. Захаров Н.П., Багдарасян А.В. Механические явления в интегральных структурах. – М.: Радио и связь, 1992. – С.72–75.
13. Пат. № 2251702, РФ. Микромеханический акселерометр. Тимошенко С.П., Рубчиц В.Г., Калугин В.В. и др. Дата подачи заявки 2004.07.02. Дата публикации 2005.05.10
14. Сайт компании Databeans – Режим доступа: <http://www.databeans.net>, свободный. – Загл. с экрана.
15. Сайт компании Freescale Inc. – Режим доступа: <http://www.freescale.com>, свободный. – Загл. с экрана.

БЕСКОНТАКТНЫЙ МЕТОД КОНТРОЛЯ МЕХАНИЧЕСКИХ ПАРАМЕТРОВ ФЕРРОМАГНИТНЫХ МАТЕРИАЛОВ

В.А. Нечаев, Н.В. Нечаева, О.И. Пирожникова

Научный руководитель – д.т.н., профессор В.Л. Ткалич

В работе дано обоснование магнитного метода контроля механических параметров. На основе проведенного исследования получены изменения коэрцитивной силы и начальной магнитной проницаемости в зависимости от величины остаточных напряжений.

Введение

Механические напряжения, приложенные к ферромагнитным материалам, изменяют их магнитные свойства. Наиболее чувствительными к изменению структуры магнитными характеристиками являются коэрцитивная сила, начальная и максимальная магнитная проницаемость и др. Для некоторых простых однофазных ферромагнетиков (Fe, Ni) установлены количественные зависимости между коэрцитивной силой, начальной проницаемостью и механическими напряжениями, приложенными и остаточными [1, 2]. В случае многофазных ферромагнетиков, каковыми являются широко используемые стали, содержащие дополнительно ряд элементов, указанные зависимости являются более сложными и до настоящего времени не установлены [5]. В этой связи представляет значительный интерес разработка метода, позволяющего установить корреляционную зависимость между магнитными и механическими свойствами сталей, с целью создания новых автоматизированных методов и средств контроля.

Основная часть

Исследования проводили на цилиндрических образцах, изготовленных из труб сталей марок: сталь 10, 35, 45, 09Г2С, 40Х, 40ГТ, бурильная Д. Магнитные характеристики (коэрцитивная сила, начальная проницаемость) определяли на баллистической установке. Механические параметры (приложенные напряжения, остаточную деформацию) измеряли на установке УММ-5. Магнитные характеристики и механические свойства определяли на одних и тех же образцах. Начальную магнитную проницаемость определяли по наклону касательной к кривой намагничивания образцов. Магнитные характеристики измеряли после снятия нагрузки [3, 4].

На рис. 1 приведены графики изменения начальной проницаемости μ_a образцов из указанных марок сталей в зависимости от величины приложенного напряжения σ .

Из графиков видно, что μ_a для всех образцов уменьшается с ростом величины напряжений. На всех экспериментальных кривых имеются горизонтальные участки, соответствующие состоянию образца, когда приложенные напряжения растяжения не превышают предела текучести. Внутри образца в этом случае оставались внутренние напряжения 2-го рода, обусловленные технологическими процессами и структурой. Параллельно на тех же образцах измеряли коэрцитивную силу H_c . Горизонтальные участки на кривых соответствуют значениям пределов текучести. Из графиков видно, что значения коэффициента k для разных материалов различны. Для исследованных материалов он изменяется от 0,042 (сталь 09Г2С) до 0,14 (сталь 40Х). Такое изменение коэффициента k можно объяснить структурным и фазовым состоянием образцов [4]. Для малоуглеродистых сталей (сталь 10 и 35) коэффициент k мало изменяется. Он равен соответственно 0,092 и 0,11. В сталях (09Г2С, 40ГТ, марка Д), содержащих легирующий элемент марганец в количествах $>1\%$, коэффициент k равен, соответственно, 0,042; 0,083 и 0,096, т.е. чувствительность коэрцитивной силы к механическим напряжениям ниже, чем у простых углеродистых сталей.

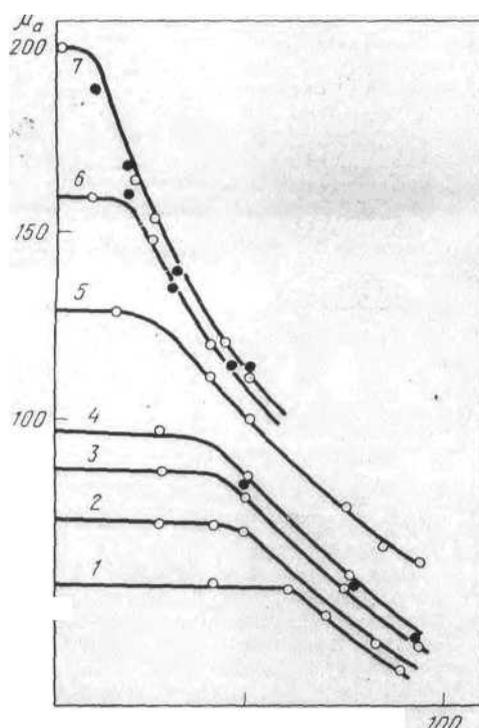


Рис. 1. Зависимость начальной магнитной проницаемости μ_0 от величины приложенных напряжений: 1 – 40Х, 2 – 40ГТ, 3 – Сталь 35, 4 – Сталь 45, 5 – Марка Д, 6 – Сталь 10, 7 – 09Г2С

Указанные различия чувствительности коэрцитивной силы к напряжениям приведенных сталей можно объяснить различным приростом плотности дефектов (дислокаций) с увеличением деформации [5]. Проведенный нами анализ коэрцитивной силы и начальной магнитной проницаемости, с одной стороны, и предела текучести указанных сталей – с другой стороны, показал, что между их значениями существует удовлетворительная корреляция.

Заключение

Таким образом, произведение коэрцитивной силы на величину начальной магнитной проницаемости представляет собой параметр, который может быть использован для контроля механических свойств сталей. Это позволит создать новые автоматизированные методы и средства контроля за состоянием неразборных узлов сложных механизмов и аппаратов, к которым затруднен доступ.

Литература

1. Франюк В.А., Харитонов А.И., Гайкович Э.Ч. // Изв. АН БССР. – Сер. физико-технических наук. – 1982. – № 1. – С 119.
2. Михеев М.Н., Горкунов Э.С, Сомова В.Н., Кутыин А.Б. // Дефектоскопия. – 1982. – № 9. – С. 66–74.
3. Вороненко Б.И. // Заводская лаборатория. – 1985. – № 1. – С. 20–25.
4. Франюк В.А., Иванькович Л.Ф., Кочломазашвили Б.Т., Дауташвили Б.Ш. // Заводская лаборатория. – 1984. – № 10. – С. 80–81.
5. Оптические методы исследования дефектов и дефектоскопия элементной базы микроэлектроники и микропроцессорной техники. Сборник научных статей. / Под редакцией д.т.н., проф. Ю.А. Гатчина и д.т.н., проф. В.Л. Ткалич. – СПб: СПбГИТМО (ТУ), 2002. – 150 с.

МАГНИТНЫЙ МЕТОД НЕРАЗРУШАЮЩЕГО КОНТРОЛЯ МЕХАНИЧЕСКИХ ХАРАКТЕРИСТИК ШИРОКОПОЛОСНЫХ СТАЛЕЙ

В.А. Нечаев, Н.В. Нечаева, О.И. Пирожникова
Научный руководитель – д.т.н., профессор В.Л. Ткалич

В работе представлена схема установки для определения механических характеристик широкополосных сталей.

Введение

На протяжении многих лет магниторезистивный (МР) эффект в ферромагнитных пленках используется при разработке датчиков магнитного поля и считывающих головок, однако внедрение тонкопленочной технологии открыло новые возможности перед разработчиками таких устройств и резко расширило круг их применений. Использование МР эффекта в тонких пленках с применением современной интегральной технологии позволяет создавать датчики и головки считывания предельно малых размеров с высокой чувствительностью. Разработкой тонкопленочных МР датчиков занимаются практически во всех промышленно развитых странах мира [3].

Известно, что механические свойства стали (в частности, результаты испытания на изгиб в холодном состоянии) тесно связаны с ее магнитными свойствами [1]. Поэтому величина коэрцитивной силы, как и пропорциональная ей сила тока размагничивания, измеряемая с помощью феррозондового коэрцитиметра, может быть использована в качестве параметра неразрушающего контроля при испытаниях на изгиб в холодном состоянии.

Основная часть

В настоящее время внедрен неразрушающий контроль изгиба широкополосной стали в технологическом потоке стана 800. Для этого разработана полуавтоматическая установка, с помощью которой, согласно ТУ I4-I-1016-74, можно измерять ток размагничивания в трех точках вдоль полосы, т.е. движение датчика имеет три степени свободы. На рис. 1 показана его схема. По рольгангу 1 движется контролируемая полоса 2. При краткой остановке рольганга (~30 с) на полосу опускается датчик коэрцитиметра 3. По поперечной направляющей раме 4 движется тележка 5 с двумя двигателями типа АОЛ-21-4 (270 Вт, 1400 об/мин; на выходном валу редуктора – 2,5 об/мин) с цепным приводом движения тележки, а также подъема и опускания датчика. С помощью пневмоцилиндра 6 направляющая рама движется относительно кронштейна 7 вдоль рольганга. Для сокращения времени замера, а, следовательно, и остановки рольганга, в коэрцитиметре применяется приставка автоматического измерения тока размагничивания [3]. Пульт управления установкой и коэрцитиметр находятся в кабине управления ножницами стана.

Коэрцитивная сила как самый структурочувствительный магнитный параметр имеет тесную связь с массовой долей химических элементов (и в частности, с повышенной массовой долей марганца) [2]. В статистическую обработку включили сталь следующих марок: Зсп производства ОХМК и ММК; 17ПС-У, 17Г1С, 17ГС, 14ГС, 09Г2, 09Г2Д, 10Г2С1 производства ОХМК; Зкп 15кп производства ММК. Стали, близкие по химическому составу, объединили в группы.

Для стали каждой марки (или группы марок) и групп толщин построили частотные распределения значения тока размагничивания. Результаты испытаний на изгиб большинства групп марок сталей, включенных в обработку, были положительные за

весь период эксперимента. Для сталей этих марок, как показал статистический анализ, ежегодно проводимый на комбинате, удовлетворительный изгиб гарантируется во всем интервале значений тока размагничивания.

Для групп марок сталей, имеющих отрицательные результаты испытаний на изгиб, значения тока размагничивания при неудовлетворительном изгибе оказались выше, чем при удовлетворительном. Для них были определены интервалы значений тока размагничивания, в которых удовлетворительный изгиб гарантируется с доверительной вероятностью 0,95. При больших значениях тока размагничивания сталь подвергают изгибу стандартным методом.

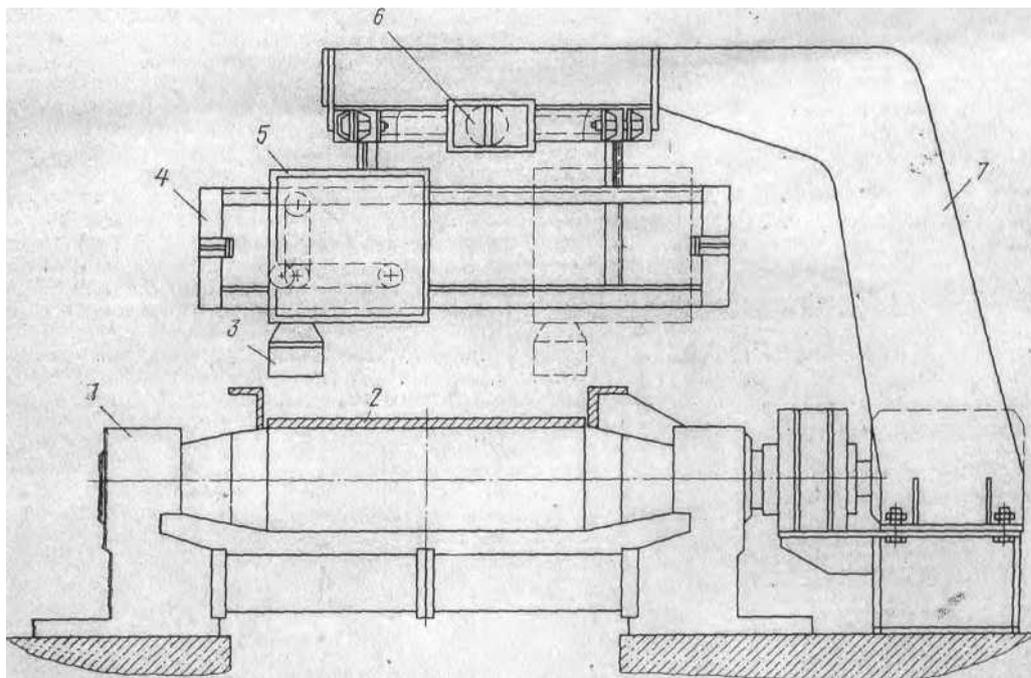


Рис. 1 Схема неразрушающего контроля механических свойств толстолистовой стали: 1 – рольганг, 2 – контролируемая полоса, 3 – датчик коэрцитиметра, 4 – рама, 5 – тележка, 6 – пневмоцилиндр, 7 – кронштейн

Заключение

Внедрение результатов работы позволило значительно сократить затраты труда, связанные с отбором проб, их раскроем, транспортировкой в ЦЗЛ, изготовлением образцов на металлообрабатывающих станках и их испытанием.

Литература

1. Григорян Г.В., Мулько Г.Н. // Дефектоскопия. – 1996. – № 1. – С. 77–80.
2. Мулько Г.Н., Григорян Г.В., Полубояринова В.Г., Бодяков А.И. // Сталь. – 1982. – № 6. – С. 76–78.
3. Кохман Л.В., Горнаков В.И. // Дефектоскопия. – 2000. – № 6. – С. 101–104.

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ДИНАМИЧЕСКИХ НАГРУЗОК НА ВЕЛИЧИНУ ДЕФОРМАЦИИ ВЕРХНЕГО СЛОЯ ФЕРРОМАГНЕТИКОВ

В.А. Нечаев, Н.В. Нечаева, О.И. Пирожникова
Научный руководитель – д.т.н., профессор В.Л. Ткалич

В работе разработана методика исследования влияния динамических нагрузок на величину деформации верхнего слоя ферромагнетиков.

Введение

Несмотря на бурное развитие науки и техники в XX веке, появление огромного количества синтетических, полимерных и композиционных материалов, конструкционные стали и в XXI веке остаются основным конструкционным материалом. В процессе эксплуатации стальных конструкций происходит естественное изменение свойств этого материала под действием окружающей среды (старение, коррозия, водородное охрупчивание и т.п.). Поэтому развитие методов оценки свойств материалов и мониторинга состояния работающего оборудования является весьма актуальной задачей.

В связи с непрерывным ростом грузонапряженности и нагрузок на оси тяжело-грузного транспорта возникает необходимость в исследованиях особенностей взаимодействия узлов при постоянно изменяющихся условиях эксплуатации. Показатели напряженно-деформированного состояния являются основой для совершенствования норм и допусков их содержания, обеспечивающих безопасность работы. Кроме того, результаты исследований силового воздействия могут быть использованы для косвенной оценки при прогнозировании интенсивности накопления остаточных деформаций и расходов на ремонты и текущее содержание.

Опытом эксплуатации и наблюдениями установлено, что около 70% всех трудовых затрат на текущее содержание приходится на выправку в профиле [1–3], что вызвано интенсивным накоплением остаточных деформаций в балластном слое. Наблюдения за опытными участками показали, что работы по выправке составляют 60–80% всех затрат рабочей силы. Основными причинами расстройств балластного слоя является неравноупругость пути по протяжению, что вызывает дополнительное динамическое воздействие.

Основная часть

Интенсивность накопления остаточных деформаций существенно зависит от величины воздействия и частоты приложения нагрузок. Экспериментальными исследованиями установлено [4, 5], что при отсутствии неровностей средние значения вертикальных сил взаимодействия в прямых практически не зависят от скорости движения. В кривых происходит закономерное перераспределение нагрузок между нитями, но в среднем они также не зависят от скорости. Средние вертикальные перемещения при изменении скорости от 40 до 100 км/ч возросли до 15%, уровень средних прогибов равен 3–4 мм. Но, если средние значения динамических вертикальных сил не зависят от скорости, то из-за наличия неровностей дисперсия с ростом скорости увеличивается. Это необходимо учитывать при прогнозировании накопления остаточных деформаций от скорости.

Что касается частоты приложения нагрузки, то в настоящее время считается, что вибрация оказывает дестабилизирующее воздействие на элементы верхнего строения. Интенсивность накопления остаточных деформаций можно определить по значению коэффициента $h\beta$, учитывающего влияние вибрационной нагрузки [6]. Анализ замерен-

ных ускорений при эксперименте показал, что следует ожидать увеличения на 30% интенсивности накопления остаточных деформаций в диапазоне скоростей 90–100 км/ч по сравнению с 80–90 км/ч.

По данным Г.М. Шахунянца [7], годовые расходы на содержание устройства пропорциональны величине

$$k = \bar{n} \cdot 2P^2 \sqrt{\frac{\bar{V}}{l}} \cdot \sqrt{T} \quad (1)$$

где \bar{n} – динамический коэффициент,

$$n_q = 1 + \alpha \bar{V} \quad (\alpha = 0,012); \quad (2)$$

P – средняя статистическая нагрузка на ось; \bar{V} – средняя скорость движения; l – среднее расстояние между осями; T – грузонапряженность в млн. ткм брутто на 1 км в год. С учетом прямой связи между расходами и затратами труда указанная зависимость остается в силе для характеристики измерения затрат труда от скорости движения.

Как было указано выше, результаты эксперимента показали, что средние вертикальные силы взаимодействия практически мало зависят от скорости движения, но значительно изменяются отклонения от них. Поэтому для учета всего силового воздействия правомерно вместо средней скорости по участку \bar{V} принимать эквивалентную скорость $V_{экр}$, которая отражает всю совокупность реализуемых скоростей. При линейной зависимости (что вполне приемлемо для технико-экономических расчетов) затрат труда на текущее содержание от скорости значение $V_{экр}$ определяется по формуле

$$V_{экр} = \bar{V} + \frac{s^2}{\bar{V}} 0,95 \text{ км/ч}, \quad (3)$$

где s – среднее квадратическое отклонение.

Формула (3) определена по методике, предложенной ЦНИИ МПС. Тогда формула (1) будет иметь вид

$$k = P^2 \sqrt{\frac{\bar{V}}{l} + \frac{s^2}{\bar{V}} \cdot 0,95} \cdot \sqrt{T} \quad (4)$$

Затраты труда на текущее содержание определяют выход элементов верхнего строения по дефектности (в том числе и по состоянию системы в плане и профиле). С учетом целей исследования были произведены статистические наблюдения и обработка информации по установлению зависимости выхода элементов и затрат труда на текущее содержание от скорости движения. При этом применялись следующие математические методы обработки данных:

- математико-статистические исчисления для установления закона и основных характеристик распределения;
- аппроксимация методом наименьших квадратов при однофакторном анализе;
- теория планирования экспериментов для качественной оценки влияния различных факторов на накопление остаточных деформаций;
- составление множественных корреляционных уравнений при многофакторном анализе;
- установление величин повреждений пути в зависимости от скорости движения поездов с учетом фактического распределения их вероятностей;
- метод статистических испытаний (Монте-Карло).

Необходимо учесть, что выход элементов по дефектности, накопление остаточных деформаций и объемов работ устанавливались как на основании теоретико-экспериментального подхода, так и обработкой фактического материала. При теоретико-экспериментальном подходе принималась эмпирическая формула с меньшими от-

клонениями расчетных и экспериментальных значений, удобная для дальнейших расчетов. А в расчетную формулу уже подставлялись данные с участка линии с определенными эксплуатационными характеристиками (скорость движения, осевая нагрузка, вид перевозимого груза).

На основании анализа фактических затрат труда P на каждый километр в течение года на 2 участках протяжением по 105 км и анализа состояния в баллах B была получена следующая зависимость расхода рабочей силы за год P_p :

$$P_p = (-0,046 + \frac{4530}{R} + 0,24 \bar{V}) \text{ чел} \times \text{год}, \quad (5)$$

По указанной зависимости расход рабочей силы был определен без учета времени на осмотр пути и проход к месту работ, затрат труда на работы, не связанные с движением. Для средних условий при повышении скорости до 30 км/ч (для фактически реализуемых скоростей движения) затраты труда, определенные по формуле (5), больше на 17%, чем при скорости 80 км/ч. Полученный результат хорошо совпадает с данными, определенными по формуле (4).

Заключение

Создание системы непрерывного контроля за работой сложного и очень важного оборудования (двигатели, турбины, реакторы) на отдаленном расстоянии требует разработки методов бесконтактной диагностики физико-механических параметров. В частности, метод определения деформации верхних слоев ферромагнетиков при динамических нагрузках – один из вариантов решения проблемы.

Литература

1. Басилов В.В., Чернышев М.А. Справочник инженера-путейца. М., Транспорт, 2002. – Т.1 – 764 с., Т.2 – 517 с.
2. Попов С.Н. Балластный слой железнодорожного пути. – М.: Транспорт, 1965. – 183 с.
3. Блохин К.А. Организация и планирование путевого хозяйства. – М.: Транспорт, 2000. – 352 с.
4. Волошко Ю.Д., Татуревич А.П., Янин В.М. Особенности вертикального воздействия на путь грузовых поездов при скоростях движения 90-100 км/ч // Труды ДИИТ. – Днепропетровск, 1997. – Вып. 188 (18).
5. Липовский Р.С., Шульман З.А. Экспериментальное исследование вертикальных сил взаимодействия между колесом и рельсом. // Труды ДИИТ. – Днепропетровск, 1994. – Вып. 151.
6. Альбрехт В.Г., Золотарский А.Ф. (под редакцией). Современные конструкции верхнего строения пути. – М.: Транспорт, 2005. – 280 с.
7. Шахунянец Г.М. Работа пути и расходы на его содержание / Вопросы пути и его содержания. – Труды МИИТ. – Вып.491. – М., 2006.

ОБЗОР МЕТОДОВ КОНТРОЛЯ ПЕРВИЧНЫХ ПАРАМЕТРОВ ОПТИЧЕСКОГО ВОЛОКНА

С.М. Иванова, О.В. Елисеев, А.С. Михайличенко, О.Л. Студеникин
Научный руководитель – к.т.н. В.Н. Фролков

Рассматриваются методы измерения характеристик оптических волокон, в частности, измерение оптической мощности, затухания и вносимых потерь. Приводится краткое описание способов контроля параметров оптического волокна.

Введение

На сегодняшний день особняком стоит проблема передачи данных. Существует большое количество способов передачи информации. Один из них – передача данных с помощью оптического волокна. Этот вариант имеет ряд неоспоримых преимуществ: большая скорость, высокое качество, перспективность, удельная дешевизна и т.д.

Создание современных волоконно-оптических линий связи требует все более высокого качества оптического волокна и жесткого контроля за его параметрами, каждый из которых обеспечивает те или иные функции системы [1].

Оптическое волокно нуждается в многократном тестировании. Производитель волокна должен убедиться в том, что волокно соответствует специфицированным характеристикам. Поскольку контроль качества осуществляется только при производстве волокна, производитель должен постоянно его тестировать с целью определения соответствия спецификации. К числу контролируемых в этих тестах параметров относятся:

- диаметр ядра волокна,
- диаметр оптической оболочки,
- апертура (NA),
- профиль показателя преломления,
- допустимая растягивающая нагрузка.

Другие тесты, выполняемые с волокнами или волоконно-оптическими кабелями, связаны с их механическими характеристиками и влиянием окружающей среды. Влияние окружающей среды оценивается по изменению затухания в волокне при экстремальных температурах, повторяющихся циклических изменениях температуры и от относительной влажности [2].

Рассмотрим основные методы измерения параметров оптических кабелей (ОК) и линейных трактов волоконно-оптических систем связи (ВОСП).

Классификация методов измерений

Методы измерения параметров ОК и линейных трактов ВОСП классифицируют по месту и цели проведения измерений, измеряемому параметру и методу измерения.

По месту и цели проведения измерений

Лабораторные измерения в научно-исследовательских институтах и заводских лабораториях отличаются высокой точностью, разнообразием применяемых методов и аппаратуры. Основная цель измерений – разработка новых типов и конструкций ОК, технологий их изготовления с целью улучшения параметров или придания им новых свойств.

Особенность *заводских измерений*, осуществляемых на производственных предприятиях в процессе изготовления ОВ – использование специализированной аппаратуры и автоматизация методов измерений. Выходными параметрами являются паспортные данные.

Измерения в процессе строительства, монтажа ВОСП и приемосдаточные испы-

тания. Число измеряемых параметров ОК и линейного тракта в процессе строительства и монтажа ограничивается обычно двумя-тремя в зависимости от типа ОК и назначения ВОСП. Приемосдаточные испытания проводятся при передаче ВОСП в эксплуатацию по заданной программе и определенной методике.

По измеряемому параметру

Измерение *оптических характеристик ОВ*, к которым относятся числовая апертура, профиль показателя преломления, диаметр модового поля и длина волны отсечки (для одномодовых ОВ), коэффициент затухания и дисперсия (полоса пропускания).

Измерение *оптических характеристик линейного тракта*, в первую очередь, измерение распределения неоднородностей, вносимых потерь и определение расстояния до места обрыва ОВ.

Измерение *конструктивных и механических характеристик ОВ*, а также их зависимостей от различных факторов, возникающих в процессе изготовления, прокладки, монтажа и эксплуатации: растяжение, изгиб, сжатие, удары и вибрация, диапазон рабочих температур, герметичность и т.д.

Измерение оптической мощности

Измерение оптической мощности осуществляется на всех стадиях производства ОВ, изготовления ОК, строительства, монтажа и эксплуатации ВОСП,

Различают измерение абсолютной и относительной мощности, В первом случае мощность выражается в единицах мощности (в ваттах), а во втором – в относительных (логарифмических) единицах по отношению к мощности 1 мВт (в децибелометрах).



Рис. 1. Схема измерителя оптической мощности

Для измерения оптической мощности используют специальный прибор – измеритель мощности (оптический ваттметр), структурная схема которого приведена на рис. 1. Сигнальный ток фотодетектора (фототок) после усилителя поступает на вход электронного осциллографа (ЭО), на экране которого можно наблюдать сигнал, пропорциональный мгновенному значению измеряемой мощности. Для получения результата измерения в цифровой форме используется аналого-цифровой преобразователь и устройство отображения. Для вывода результата в логарифмических единицах используется логарифматор [5].

Для расширения функциональных возможностей измерителя оптической мощности он обычно дополняется источником излучения, что позволяет использовать его для измерения потерь (затухания). Прибор, в котором используется согласованная пара источник излучения–измеритель оптической мощности, называется оптическим тестером или измерителем оптических потерь. Очевидно, что, выполняя измерение оптической мощности или уровня мощности на входе P_1 , p_1 и выходе P_2 , p_2 оптического кабеля, можно определить его затухание и коэффициент затухания:

$$a = 10 \lg \frac{P_1}{P_2} = p_1 - p_2, \text{ дБ}; \alpha = \frac{a}{l}, \text{ дБ/км}, \quad (1)$$

где l – длина оптического волокна, км.

В соответствии с ГОСТ 26814-84 и Рекомендациями МСЭ-Т G 651 и G 652 разли-

чают три основных метода измерения затухания: метод обрыва, метод вносимых потерь и метод обратного рассеяния [7].

Измерение затухания методом обрыва

Метод обрыва основан на сравнении значений оптической мощности, измеренной на выходах длинного и короткого отрезков ОВ (порядка 2 м), обломанного от его начала. Метод обеспечивает высокую точность измерений затухания, но его использование в полевых условиях и при техническом обслуживании затруднительно. Поэтому его обычно применяют при производстве ОВ и ОК.

Упрощенная схема установки для измерения по методу обрыва представлена на рис. 2. Излучение оптического источника проходит через смеситель мод, фильтр вытекающих мод и через устройство ввода поступает в ОВ. Выходную оптическую мощность ОВ или короткого отрезка определяют измерителем оптической мощности.

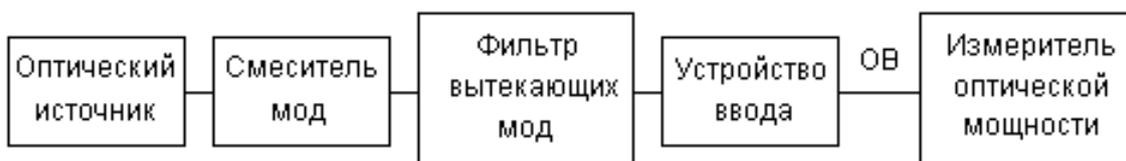


Рис. 2. Схема установки для измерения затухания по методу обрыва

Процесс измерения заключается в следующем. Оба конца измеряемого оптического волокна освобождают от защитных покрытий и обрабатывают так, чтобы они были ровными, перпендикулярными оси и не имели сколов и трещин. С помощью устройства ввода проводят юстировку и фиксацию входного торца и измеряют значение выходной мощности P_2 . Затем, не изменяя положения оптического волокна в устройстве ввода, обламывают измеряемое оптическое волокно на расстоянии 1,5–2 м от входного торца. Обрабатывают конец короткого отрезка и измеряют значение выходной мощности. Эту мощность принимают равной мощности P_1 на входе измеряемого оптического волокна. По формулам (1) определяют затухание и коэффициент затухания.

Важным является выбор источника оптического излучения. Мощность P_1 и длина волны излучения λ должны быть высокостабильными. Кроме того, P_1 должна быть достаточно большой, а спектр излучения – узким. Это увеличивает диапазон измерений затухания и обеспечивает единство измерений разными приборами.

В многомодовых оптических волокнах измеряют некоторое среднее затухание, обусловленное всеми направляемыми модами. Поскольку направляемые моды изображаются лучами, проходящими различными путями вдоль сердцевины оптического волокна, затухание их различно. Очевидно, что измеряемое затухание зависит от степени взаимодействия мод, т.е. от распределения передаваемой энергии между ними. Равновесное распределение устанавливается на длине установившейся связи L_v , которая превышает 5 км. Поэтому, чтобы результаты измерения затухания в многомодовых оптических волокнах не зависели от длины (особенно при малых длинах), необходимо искусственно обеспечить равновесное распределение энергии между модами на всем протяжении измеряемого оптического волокна. С этой целью обычно используют дополнительные оптические волокна с длиной $L > L_v$ или специальные смесители мод, которые представляют собой стержень диаметром 15–40 мм и длиной порядка 20 мм с несколькими витками ОВ, намотанными с небольшим натягом.

При измерении затухания на коротких оптических волокнах возникает погрешность за счет вытекающих мод, которые, хоть и затухают значительно быстрее направляемых мод, могут распространяться на большие расстояния. На выходе отрезка опти-

ческого волокна измеритель оптической мощности регистрирует дополнительную мощность вытекающих мод. Эта погрешность может быть большой в одномодовых оптических волокнах. Для устранения вытекающих мод обычно используют фильтр вытекающих мод, который представляет собой отрезок оптического волокна, освобожденный от защитных покрытий и помещенный в иммерсионную жидкость. Если показатель преломления иммерсионной жидкости $n \geq n_2$, где n_2 – показатель преломления оболочки оптического волокна, вытекающие моды покидают оптические волокна и рассеивают свою энергию в иммерсионной жидкости.

Измерение затухания методом вносимых потерь

Метод вносимых потерь менее точен, чем предыдущий, но не требует обрыва оптического волокна и поэтому обычно используется в полевых условиях. Для измерения затухания используется тестер, у которого измеритель оптической мощности и источник излучения оборудованы оптическими разъёмными соединителями. Разъёмными соединителями также армированы измеряемое оптическое волокно и оптический поводок (ОП) (эталонное оптическое волокно) [9].

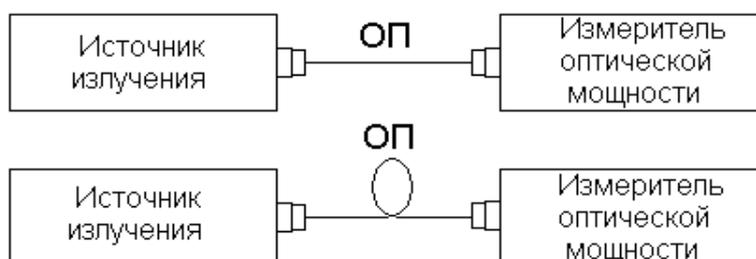


Рис. 3. Схема установки для измерения затухания методом вносимых потерь

Схема установки для измерения затухания представлена на рис. 3. Выполняют два измерения. Сначала определяют оптическую мощность на выходе оптического поводка P_1 . Затем между источником излучения и измерителем мощности подключают измеряемое оптическое волокно и определяют мощность на его выходе P_2 . Затухание и коэффициент затухания оптического волокна при коротком поводке определяют по формулам (1). При длинном поводке в результаты измерения входит заданное затухание оптического поводка a_n :

$$a = \alpha L = 10 \lg \frac{P_1}{P_2} + a_n, \text{ дБ.} \quad (2)$$

В условиях эксплуатации при измерениях в проложенном оптическом кабеле, когда вход и выход оптического волокна разнесены на значительное расстояние, при определении вносимого затухания возникает необходимость использования двух измерителей оптической мощности, показания которых предварительно сравнивают по одному источнику оптического излучения.

Кроме погрешности, вызванной нестабильностью параметров оптического источника и разъёмных соединителей, на результаты измерений существенно влияют изгибы оптического волокна. Однако в целом точность измерений по данному методу оказывается вполне приемлемой для практических целей.

Измерение вносимых потерь оптических элементов

Вносимые потери таких оптических элементов, как оптические соединители, фильтры, компенсаторы дисперсии, разветвители, мультиплексоры и т.п., при вводе их в

линейный тракт ВОСП существенно влияют на энергетический потенциал системы. При измерении вносимых потерь приходится учитывать, что разные измеряемые элементы (ИЭ) отличаются типом оптических входов и выходов. Как правило, они представляют собой гибкие одноволоконные шнуры, армированные разъемными соединителями. В этом случае измерения проводят на установке, схема которой представлена на рис. 3.

Процесс измерения состоит из двух последовательных этапов. На первом этапе измеряют мощность на входе ИЭ. Для этого источник излучения и измеритель мощности соединяют между собой оптическим поводком, желательного того же типа, что и шнуры ИЭ. На втором этапе вместо оптического поводка включают ИЭ и измеряют мощность на его выходе. Величину вносимых потерь определяют как отношение измеренных значений мощности, выраженное в децибелах. При таком методе измерений, как и при методе, описанном в предыдущем пункте, в результате входит затухание оптического поводка. Чтобы избежать этого, можно на втором этапе включить ИЭ между поводком и измерителем мощности.

Метод обратного рассеяния

Наиболее универсальным и информативным методом измерения параметров оптического волокна является метод обратного рассеяния. В настоящее время он получил наибольшее распространение при строительном-монтажных работах и в процессе эксплуатации ВОСП. Приборы, основанные на методе обратного рассеяния, называют оптическими рефлектометрами [8]. Упрощенная структурная схема и принцип действия оптического рефлектометра представлены на рис. 4.



Рис. 4. Упрощенная структурная схема и принцип действия оптического рефлектометра

Измерение конструктивных и механических параметров ОВ

Конструктивные и механические параметры ОВ и ОК имеют первостепенное значение для обеспечения высокой надежности и требуемых параметров передачи линейных трактов ВОСП.

Измерение конструктивных параметров ОВ состоит в определении диаметров сердцевин и оболочек на концах ОВ и их изменений по его длине. Измерение изменения диаметра ОВ по его длине осуществляется в процессе вытяжки ОВ в датчике, где используется лазерная система измерения диаметра. Измерение диаметров сердцевин и оболочек ОВ и отклонение сечения от концентрической формы на концах производят визуальным путем с применением микроскопа с измерительной сеткой. Эти измерения часто выполняют при монтаже неразъемных и разъемных соединителей [3, 6].

В процессе изготовления, прокладки, монтажа и эксплуатации ОК подвергаются растяжениям, изгибам, вибрационным и ударным нагрузкам, сжатиям и другим воздействиям статического механического характера. Стойкость ОВ к этим видам нагрузок, пригодность ОК различных типов для строительства и эксплуатации в требуемых условиях определяют путем соответствующих испытаний и измерений механических параметров опытных образцов на специальных стендах.

Заключение

Измерения первичных параметров оптоволокна можно провести с помощью специальной установки для измерения возвратных потерь.

Тесты проводятся в течение и после прокладки волоконно-оптической линии. Такие тесты гарантируют, что установленная система соответствует требованиям производительности. Затухание в соединителях, например, не должно превышать определенной величины. В ходе установки каждый соединитель должен быть протестирован.

Таким образом, целью данной статьи было краткое описание способов контроля параметров оптического волокна.

Литература

1. J. Beller. OTDRs and backscatter measurements // Fiber Optics Test and Measurements. Edited by D. Derricson. – New Jersey, Prentice Hall PTR, 1998. – 434 p.
2. Gardner W.B., Nagel S.R. // The Bell Syst. Tech. J. – 1981. – Vol. 60. – №6. – P. 859.
3. Берлин Б.З. Волоконно-оптические системы связи на ГТС. – М., 2004. – 386 с.
4. Гроднев И.И. Линии связи. – М., 1995. – 227 с.
5. Воронцов А.С. Оптические кабели связи российского производства. – М., 2003. – 321 с.
6. Иванов А.В. Волоконная оптика. – М.: Компания Сайрус системс, 1999. – 342 с.
7. Исаев С.К. Физика волоконно-оптических устройств. – М.: МГУ, 1986. – 219 с.
8. Корниенко Л.С., Наний О.Е. Физика лазеров. – М., МГУ, 1996. – 334 с.
9. Стерлинг Д. Техническое руководство по волоконной оптике. – М.: Лори, 1998. – 542 с.

**УТОЧНЕНИЕ РАБОЧЕГО ЧАСТОТНОГО ДИАПАЗОНА
АППАРАТНО-ПРОГРАММНОГО КОМПЛЕКСА СТАП****А.Ю. Гришенцев****Научный руководитель – д.т.н., профессор К.Г. Коротков**

Рассмотрены вопросы электропроводящих особенностей биологических тканей в выделенном частотном диапазоне 0–10 МГц. Выделен частотный диапазон, представляющий наибольший интерес с точки зрения изучения особенностей и свойств живых биологических тканей.

Введение

Выделение наиболее информативного частотного диапазона при измерении полного импеданса биологических тканей для исследования их свойств позволяет производить более детальное изучение особенностей данных тканей.

Задачи и методики исследований

Задачами исследования являлось выделение наиболее информативного частотного диапазона импеданса биологических тканей при стимуляции низким напряжением.

В процессе работы был изготовлен макет прибора, позволяющего сделать необходимые исследования.

В ходе исследования были произведены измерения и предварительный анализ амплитудно-частотной характеристики (АЧХ) сопротивления биологических тканей и объектов, некоторых жидкостей, а также исследована динамика поведения АЧХ в зависимости от времени и воздействующих факторов.

Результаты исследований

Рассмотрим АЧХ зависимости полного импеданса от частоты, полученные опытным путем при помощи разработанного макета прибора СТАП, для стандартной RC-модели биологических тканей (рис. 1) [1, 2], и АЧХ, смоделированную в вычислительном комплексе MicroCap. Значения номиналов сопротивлений и конденсатора: $R1=27$ кОм; $R2=47$ Ом; $C=3$ нФ.

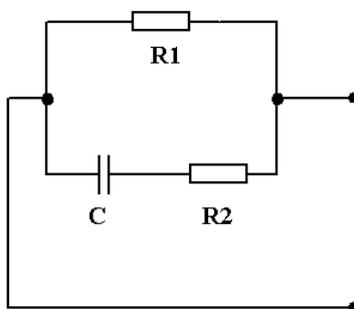


Рис. 1. RC модель биологических тканей

При размещении двух кривых АЧХ на одном графике (рис. 2) было произведено перемасштабирование данных полученных при помощи прибора, так как в процессе

вывода данных в ПК из блока измерения производится их оцифровка в относительные величины. Видно, что результаты достаточно хорошо совпадают, что говорит о корректной работе экспериментальной модели прибора.

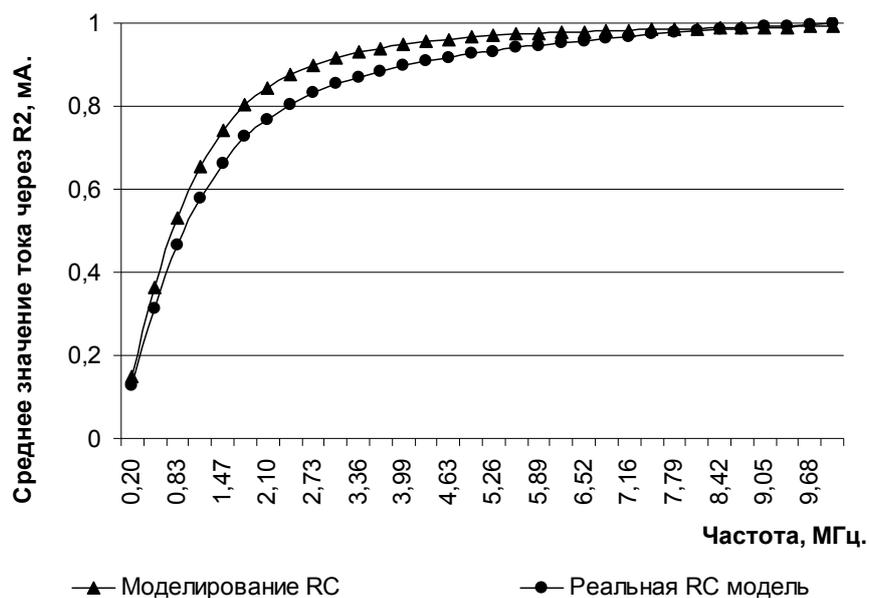


Рис. 2. Сопоставление результатов моделирования с реальными измерениями

Некоторые практические результаты экспериментов полученных при помощи АПК СТАП приведены на графике (рис. 3).

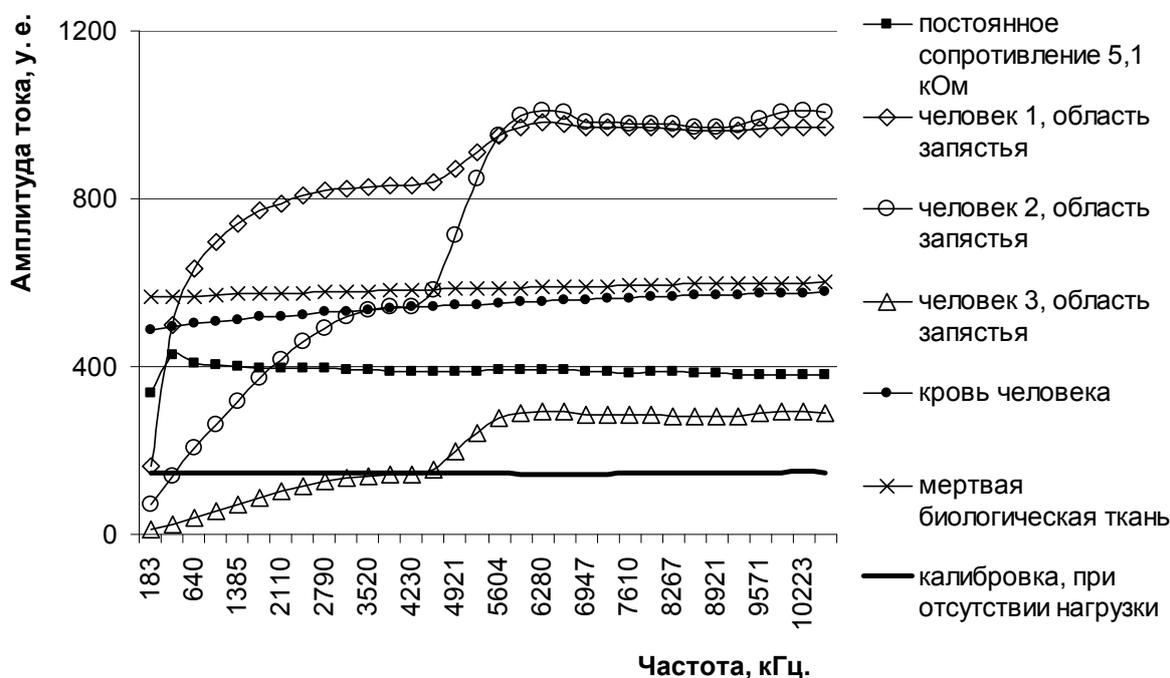


Рис. 3. Результаты измерений

Нагрузка в виде постоянного сопротивления и калибровка (при отсутствии нагрузки на измерительных электродах) дает практически горизонтальную прямую, что и следовало ожидать. Это измерение в то же время показывает, что межэлектродная ем-

кость и емкость коаксиального кабеля, по которому подавался электрический потенциал к объекту измерений, практически не влияет на результаты измерений в выбранном частотном диапазоне (до 10 МГц). Характерный вид имеет график для некротизированной ткани (в качестве объекта измерений использовалось свиная тушка). Кровь (исследовалась спустя не более 2 мин. после забора) также, несмотря на наличие в ней мембранных структур, незначительно проявляет емкостные свойства в силу больших расстояний между кровяными тельцами, плавающими в плазме. Измерения тканей человека показывают ярко выраженное проявление емкостных свойств плотных, живых биологических тканей. Наиболее интересны для анализа области: резкого роста 0–3,5 МГц, далее «ступенька» 3,5–4,5 МГц; следующая область резкого роста 3,5–6 МГц; всплеск в области около 6,2 МГц; далее незначительный спад к 7 МГц и выход на практически прямую, близкую к горизонтальной. Интересен график, полученный для 3-го человека, он имеет значительно меньшую амплитуду для всего диапазона частот по сравнению с данными для 1-го и 2-го человека. Это можно объяснить высоким сопротивлением кожи (значительная толщина, сухость).

Важно отметить, что, несмотря на различие всех АЧХ для разных людей, полученных в этих и других экспериментальных данных, основные тенденции, характерные для указанных областей частот, сохраняются. Это объясняется совпадением частот диэлектрорефа и электровращения, т.е. одинаковостью строения мембран, для биологических тканей различных людей, но интенсивность проявления этих свойств различна, как различно состояние тканей у разных людей и одного человека при различных условиях.

Обсуждение

Предпосылками к выбору рабочего частотного диапазона послужили некоторые теоретические и экспериментальные данные.

На рис. 4 показана частотная зависимость диэлектрической проницаемости, в которой выделяются области α -, β - и γ -дисперсии.

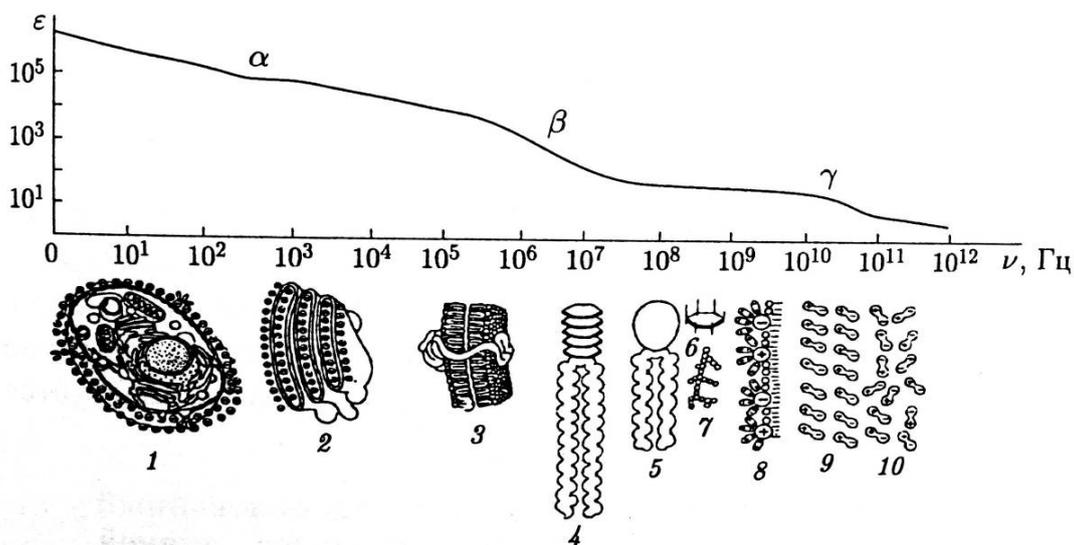


Рис. 4. Частотная зависимость диэлектрической проницаемости

Наличие областей дисперсии связано с различными механизмами поляризации тканей в различных частотных диапазонах действующего внешнего электромагнитного поля. Каждый из механизмов поляризации характеризуется своей частотой, вблизи которой запаздывание смещения (поворота) различных элементов биологических тканей минимально.

Рассмотрим механизмы возникновения областей дисперсии.

α-дисперсия обусловлена поляризацией целых клеток (1, 2) в результате диффузии ионов, что требует относительно большого времени, поэтому данный механизм проявляется при действии электрического поля (0,1–10 кГц); в этой области емкостное сопротивление мембран очень высоко и преобладают токи, протекающие через растворы электролитов, окружающие фрагменты мембран.

β-дисперсия обусловлена структурной поляризацией клеточных мембран (3), в которых участвуют белковые макромолекулы (4), а на их верхней границе – глобулярные водорастворимые белки (5), фосфолипиды (6, 7) и мельчайшие субклеточные структуры (8). Приведем частоту вращательного движения некоторых мембранных молекул: фосфолипидов $(1-2) \cdot 10^8$ Гц, родопсин около $5,5 \cdot 10^5$ Гц, цитохромоксидаза $1,5 \cdot 10^3$ Гц [3]. Данный механизм проявляется при действии электромагнитного поля с частотой 1 – 10 МГц.

γ-дисперсия обусловлена процессами ориентационной поляризации молекул (9, 10) свободной и связанной воды (частота около 20 ГГц), а также низкомолекулярных веществ типа сахаров и аминокислот (1–10 ГГц).

Таким образом, достаточно интересными для исследований и получения практических результатов являются области *α*- и *β*-дисперсии, частотный диапазон до 10 МГц, именно в этих областях происходит характерные только для живых тканей резонансные эффекты. Особенно интересна область *β*-дисперсии, в которой проявляется ответная реакция структур клеточных мембран на воздействие внешним переменным электромагнитным полем. Также представляет интерес динамика изменения полного импеданса при переходе из области частот порядка сотен кГц в область единиц МГц, связанная с емкостными свойствами живых клеточных структур.

В области частот до десятков МГц эквивалентная глубина проникновения в ткани достаточно высока (5–10 см), такое обстоятельство позволяет производить обследование не только кожных и поверхностных подкожных слоев, но и более глубоко залегающих тканей. В пользу ограничения измерений частотой 10 МГц говорит немало-важный фактор конструкторского исполнения, в области более высоких частот СВЧ и КВЧ диапазона необходимо использовать совершенно другие принципы исследований.

Заключение

На основании предварительных исследований можно сделать вывод, что наибольший интерес с точки зрения исследования состояния биологических тканей посредством изучения ее полного импеданса представляет область 0,2–7,5 МГц.

Изучение динамики изменения АЧХ полного импеданса биологических тканей в выделенном частотном диапазоне позволяет неинвазивно изучать изменения состояние биологических тканей на клеточном уровне.

Литература

1. Федорова В.Н., Степанова Л.А. Краткий курс медицинской биологии с элементами реабилитологии. Лекции и семинары: Учебное пособие. – М.: ФИЗМАТЛИТ, 2005. – 624 с.: ил.
2. Смолянинов В.В. Математическое моделирование биологических тканей. – М.: Наука, 1980. – 368 с.: ил.
3. Самойлов В.О. Медицинская биофизика: учебник для вузов. – 2-е изд., испр. и доп. – СПб.: СпецЛит, 2007. – 560 с.: ил.

**ФОРМИРОВАНИЕ СТРУКТУРНО-ФУНКЦИОНАЛЬНОЙ СВЯЗИ
«ЗАКАЗЧИК – ПОЛЬЗОВАТЕЛЬСКАЯ СЕТЬ» СУ КМП
В СИСТЕМЕ ОМС****Е.О. Цингер****Научный руководитель – к.т.н., доцент З.Г. Симоненко**

Рассмотрено формирование структурно-функциональной связи «заказчик – пользовательская сеть» системы управления качеством медицинской помощи (далее – СУ КМП) в системе обязательного медицинского страхования (далее – ОМС) при разработке программных средств для сбора, хранения и статистической обработки данных СУ КМП.

Введение

Совместно с органами управления здравоохранения Санкт-Петербурга Территориальным фондом обязательного медицинского страхования в районах Санкт-Петербурга создается и внедряется СУ КМП в системе ОМС. Субъектами СУ КМП являются: Правительство Санкт-Петербурга, Администрации районов Санкт-Петербурга, Территориальный фонд ОМС Санкт-Петербурга (далее – ТФ ОМС), амбулаторно-поликлинические учреждения (далее – АПУ), страховые медицинские организации (далее – СМО), а также конечные пользователи – граждане РФ.

Для управления КМП в районах Санкт-Петербурга, а также для принятия надлежащих управленческих решений необходимо:

- организовать мониторинг качественных характеристик АПУ – укомплектованность кадрового состава врачами-специалистами с учетом их квалификации, наличие современной материально-технической базы, наличие нормативно-правовой базы и т.д.;
- производить статистическую и аналитическую обработку данных АПУ.

В условиях большого объема информации и сложного сочетания административных и договорных отношений между субъектами СУ КМП обмен информацией осложнен. Организовать обмен и обработку этой информации для дальнейшего использования наиболее целесообразно с помощью современных информационных технологий.

Для создания информационной системы [1], которая могла бы обеспечить такой обмен информацией, удобный для каждого из субъектов, необходимо рассмотреть структурно-функциональные связи между ними.

Структурные связи

Как было сказано во введении, некоторые субъекты не имеют иерархической (административной) связи, но при этом одни из них подотчетны другим не в силу административной подчиненности, а по условиям заключения договоров. На рис. 1 приведена схема, которая в некоторой степени раскрывает эти связи. Пунктирной линией отмечены связи, носящие только договорной характер.

Правительство Санкт-Петербурга, в том числе Губернатор Санкт-Петербурга, принимают административные решения через издание распоряжений и приказов Правительства относительно деятельности Администраций районов Санкт-Петербурга, РУЗ и АПУ [2]. Между остальными субъектами существуют различные договорные отношения. Между СМО и АПУ заключаются договоры на лечебно-профилактическую

помощь в системе ОМС, между СМО и ТФ ОМС – договоры о финансировании в системе ОМС, между ТФ ОМС и Администрации районов – соглашения о совместной деятельности.

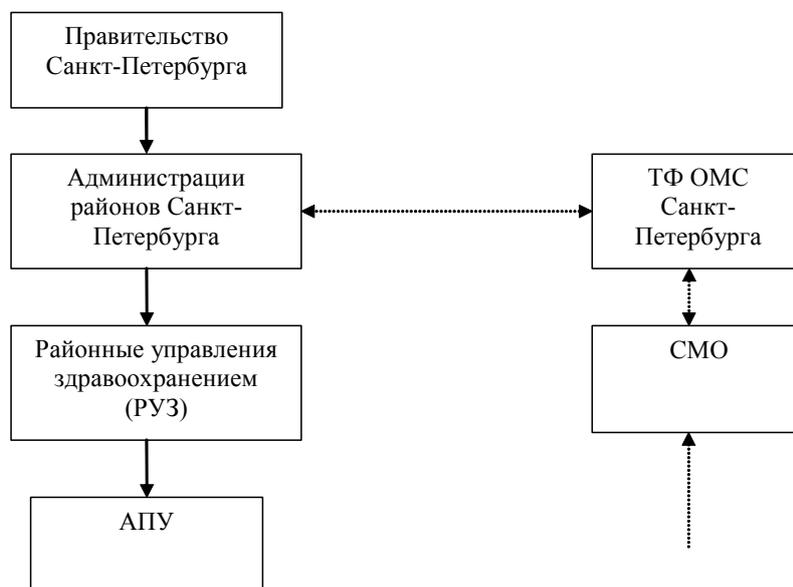


Рис. 1. Структурные связи между субъектами-участниками системы управления качеством медицинской помощи в районах Санкт-Петербурга

Вы могли заметить, что в приведенной схеме отсутствует один из субъектов, а именно граждане РФ. Это объясняется тем, что граждане в данном случае не участвуют непосредственно в административных или договорных отношениях с другими субъектами. Гражданин получает информацию, таким образом реализуя свое право на выбор АПУ, может потребовать отчета у других субъектов, а также участвует в формировании одной из составляющих качества медицинской помощи – удовлетворенность потребителя (пациента) медицинским обслуживанием¹ [3].

Обязательства о предоставлении информации должны быть включены в соглашения Администраций районов и ТФ ОМС Санкт-Петербурга. По условиям этого соглашения ТФ ОМС будет являться ответственным субъектом за мониторинг и первичную статистическую обработку мониторируемых данных. Другие субъекты, включая граждан РФ, в пределах своей компетенции будут пользоваться результатами обработанной и опубликованной информации.

Функциональные связи

На совещании с участием губернатора Санкт-Петербурга В.И. Матвиенко было принято решение о создании единой системы управления качеством медицинской помощи в Санкт-Петербурге [4]. В рамках этого решения должна быть создана информационная система, призванная обеспечить сбор, хранение и обработку данных СУ КМП в районах Санкт-Петербурга.

¹ Согласно рекомендации экспертной группы Европейского регионального бюро Всемирной организации здравоохранения (ВОЗ) «Принципы обеспечения качества», существует четыре компонента КМП: соблюдение медицинских технологий (выполнение врачебного процесса), риск травм и заболеваний в результате медицинского вмешательства, оптимальность использования ресурсов здравоохранения и удовлетворенность потребителей медицинской помощи.

Таким образом, заказчиком данной системы является Правительство Санкт-Петербурга, исполнителями – ТФ ОМС Санкт-Петербурга совместно с Администрациями районов, конечным же пользователями результатов этой работы станут все указанные субъекты СУ КМП, в том числе Правительство Санкт-Петербурга в лице Комитета по здравоохранению (рис. 2).

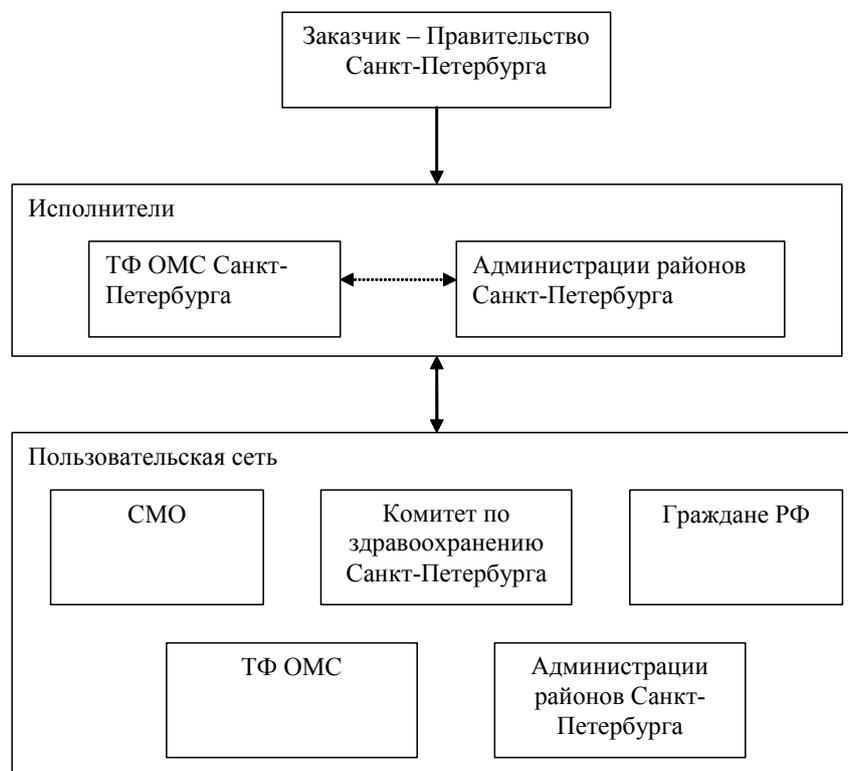


Рис. 2. Функциональные связи между субъектами системы управления качеством медицинской помощи в районах Санкт-Петербурга

Субъекты пользовательской сети взаимодействуют друг с другом посредством разного рода взаимоотношений (совместные совещания, письменные обращения граждан, дополнительные соглашения, издание приказов и распоряжений и т.д.). Результатом этих взаимоотношений будет обратная связь между исполнителями и пользователями. Обратные связи позволят обеспечить возможность для совершенствования информационной системы, как на этапе создания, так и на этапе эксплуатации.

Информационная система может стать источником информации для субъектов СУ КМП для улучшения КМП, отслеживая материально-техническую базу медицинских учреждений, обеспеченность АПУ медицинским персоналом и узкими специалистами, нормативно-правовую базу АПУ, проведение и возможности для проведения экспертиз КМП по утвержденной методике. Отслеживая эти параметры, субъекты СУ КМП получают возможность формирования индивидуальных и общих программ по улучшению КМП.

Далее будут описаны взаимоотношения (в данном случае – обмен информацией) субъектов СУ КМП на этапе эксплуатации информационной системы.

Обмен информацией

Между субъектами предусмотрен обмен данными в информационной системе СУ КМП (рис. 3).

В процессе сбора данных около 350 АПУ Санкт-Петербурга предоставляют свои данные в Администрации районов. В свою очередь, Администрации 19 районов Санкт-Петербурга с помощью предлагаемой информационной системы передают собранную информацию в ТФ ОМС. В ТФ ОМС эта информация обобщается в единую базу данных и подвергается статистической обработке. Опубликованными материалами могут воспользоваться каждый из субъектов СУ КМП.

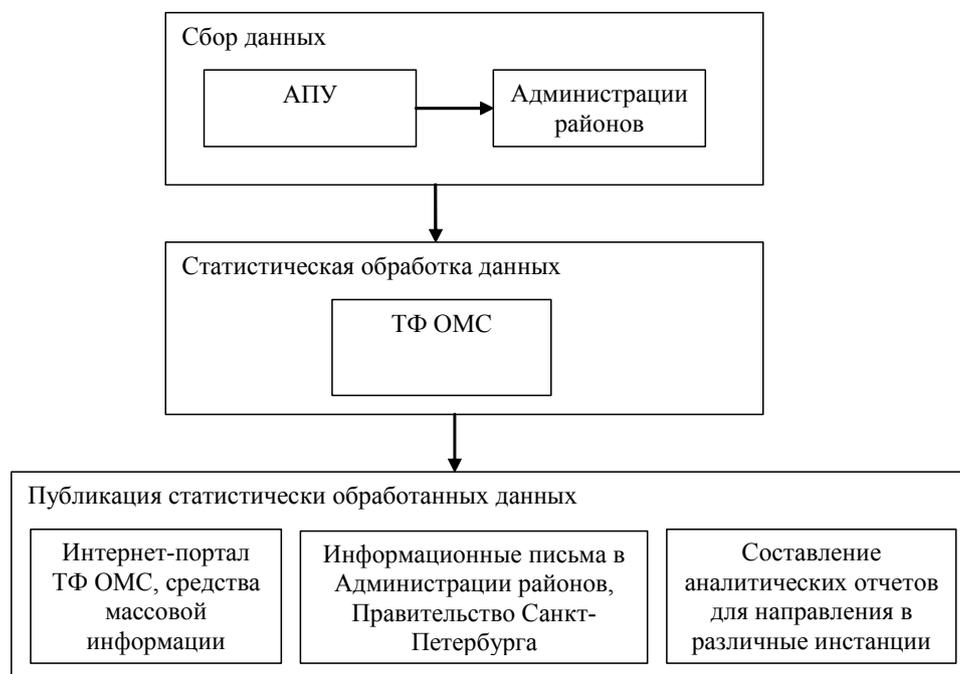


Рис. 3. Информационный обмен между субъектами СУ КМП

Заключение

Рассмотренные структурно-функциональные связи позволили оценить возможность создания системы организованного обмена информацией в части управления КМП между субъектами-участниками системы ОМС в Санкт-Петербурге.

На сегодняшний момент совершенно очевидна необходимость системного подхода к проблеме КМП в АПУ районов, так как это позволит различным органам управления здравоохранением принимать решения по организации качества и доступности медицинской помощи, в том числе в АПУ (около 80% медицинской помощи оказывается в амбулаторно-поликлинических условиях). Для разрешения этой задачи необходимо организовать обмен данными СУ КМП между всеми участниками системы ОМС. Осуществить такой обмен позволит создание информационной системы, которая позволит организовать:

- принятие управленческих решений по улучшению КМП Администрациями районов и Комитетом по здравоохранению Санкт-Петербурга;
- регулирование тарифов оказания медицинской помощи в АПУ Комитетом по здравоохранению по данным статистической обработки;
- регулирование договорных отношений с АПУ страховыми медицинскими организациями;
- реализацию права выбора АПУ застрахованным гражданином на основе опубликованных данных статистической обработки полученной информации.

Таким образом, между субъектами-участниками системы ОМС необходим своевременный и организованный обмен информацией, который возможно осуществить

через создание информационной системы по сбору, хранению и статистической обработке данных СУ КМП в административных районах Санкт-Петербурга.

Литература

1. Зиндер Е. Что такое «информационная система». – Режим доступа: <http://www.osp.ru/cio/2002/06/172182/>, своб.
2. Устав Санкт-Петербурга (принят Законодательным Собранием Санкт-Петербурга 14 января 1998 года, с изменениями на 13 декабря 2007 года).
3. Принципы обеспечения качества. Отчет о совещании ВОЗ. Барселона, 17-19 мая 1983 г./ Пер. с англ. ВОЗ. Европейское региональное бюро. Отчеты и исследования. – М.: Медицина, 1991(94). – 27 с.
4. Протокол совещания с участием Губернатора Санкт-Петербурга «О работе страховых медицинских организаций в системе обязательного медицинского страхования Санкт-Петербурга» (протокол № 45-в от 05.04.2007).

ПОДДЕРЖАНИЕ КАЧЕСТВА WEB-ПРИЛОЖЕНИЯ, НАПИСАННОГО НА ЯЗЫКЕ PHP

Д.Г. Юдин

Научный руководитель – к.т.н., доцент Б.А. Крылов

В статье рассматриваются методы повышения и поддержания на должном уровне качества web-приложения. Так же приведено сравнение различных сред для создания тестов. Статья будет полезна тем, кто только собирается вводить тестирование своих приложений и не знает, с чего начать.

Введение

Язык PHP давно уже перестали считать каким-то несерьезным и годным только для написания домашних страничек или простеньких форумов. Сегодня PHP определенно лидирует в области построения web-ориентированных проектов, будь то информационные сайты, Интернет-магазины, игры, системы документооборота и даже системы, управляющие оборудованием.

Появились профессионалы, которые успешно применяют этот язык для решения бизнес-задач. При создании продукта обычно заключается договор на поддержку. Поддержка же состоит в исправлении мелких недочетов, выявленных после окончания гарантийного периода, а также в текущих мелких и не очень мелких изменениях. Учитывая, что любое изменение может внести ошибку в приложение, возникает необходимость проверки работоспособности программы после изменения. На это требуется некоторое время. Часто применяется практика поддержки готового продукта другим программистом (не тем, который создавал данное приложение), который не вполне представляет, какие части приложения может затронуть данное изменение. Тогда приходится либо проверять все приложение целиком, а это не слишком удачное применение рабочего времени программиста, либо надеяться на удачу, что так же неразумно.

При поддержке какого-либо продукта главная задача – минимизировать затраты на выполнение работ. Имеется в виду время, потраченное на введение изменения. Для этого работа должна быть сделана «с первого раза». В статье рассмотрены некоторые приемы и методики снижения количества ошибок и уменьшения времени на их обнаружение, а также способы поддержания качества web-приложения.

Приемы, способствующие уменьшению ошибок при написании кода [1]

Первое, что необходимо сделать, – это включить отображение всех ошибок, предупреждений различных уровней. К ошибкам относятся ошибки парсера, синтаксические ошибки. К предупреждениям – использование неинициализированных переменных, обращения к несуществующему элементу массива и т. п.

Отображение всех «неправильных» ситуаций стимулирует к написанию более правильного с точки зрения стандарта языка кода. Таким образом, уменьшается количество программных ошибок, возникающих в непредвиденных ситуациях.

Следующий прием – довольно спорный, хотя также способствует уменьшению количества ошибок. В языке PHP допускается присваивание значения переменной внутри условного оператора. Таким образом, довольно часто из-за опечатки возникает вот такая конструкция:

```
if( $param = 10 ) {  
    ...  
}
```

Такую опечатку искать достаточно сложно, так как никаких предупреждений не возникает. Программа проходит по ошибочному пути, а различие между = и == в глаза не бросается. В этом случае предлагается менять местами операнды сравнения.

```
if( 10 == $param ) {  
    ...  
}
```

Тогда при опечатке парсер выдаст сообщение об ошибке, так как константе нельзя присвоить значение. Впрочем, как упоминалось ранее, такая конструкция вызывает много споров. Ведь она несколько нелогична, все же мы проверяем значение \$param. Масла в огонь также подливают высказывания, что в «нормальных» языках присваивание внутри условия недопустимо.

Сопровождение приложения

Как сказано выше, обычно заключается договор на поддержку web-приложения. Как правило, этот договор срабатывает при каких-либо непредвиденных обстоятельствах вроде атаки на сайт. Поэтому лучше быть готовым до того, как произойдет что-нибудь плохое. Для этого хорошей практикой будет постоянное отслеживание работы приложения в критических местах и отправка уведомлений, если случилось нечто, выходящее за рамки ожиданий. Например, следует отслеживать ошибки выполнения запросов к базе данных и/или значения переданных в функцию параметров. Также полезно проверять вывод в браузер. Например, в ExpertCMS для имен переменных в шаблонизаторе используется конструкция {имя переменной}. Наличие такой строки в выводе может означать, что какой-то из модулей отработал неправильно либо вообще не отработал.

В языке PHP есть встроенный механизм для добавления проверок в код. Это так называемые assert-проверки. Суть их состоит в том, что в код программы добавляются проверочные конструкции, которые можно централизованно включать/выключать. С помощью них удобно проверять, например, параметры, переданные в функцию. Кроме того, можно определить обработчик ошибок, в котором написать, что с этой ошибкой сделать. Например, в течение гарантийного периода полезно оставить включенными assert-проверки, а в обработчике ошибок сделать отправку сообщений об ошибках программисту на почту. В таком случае не придется допрашивать клиента, что он сделал, чтобы ошибка возникла. К тому же, можно исправить ошибку до того, как ее последствия зашли слишком далеко. Например, злоумышленник, взламывающий сайт через sql-injection, скорее всего своими попытками сгенерирует ошибку в запросе к базе данных.

Приемочное тестирование

Под приемочными тестами понимают тесты, которые проводятся в момент, когда разработчики сообщают о конце разработки. Этими тестами проверяется соответствие поведения готового продукта тому поведению, которое описано в техническом задании и в userstory. Такое тестирование, как правило, происходит по методике черного ящика, т.е. имитируются действия обычного пользователя программы. Как правило, это тестирование работы интерфейса. Оно связано с взаимодействием с элементами управления (кнопки, переключатели, флаги) или вводом. Очевидно, что проводить такое тестирование раз от раза ручным способом довольно долго, да и вероятность что-либо пропустить достаточно высока. Поэтому эти тесты необходимо автоматизировать.

Для web-приложений тестирование интерфейса можно осуществлять с помощью пакета программ Selenium. Данный инструмент появился относительно недавно. Первая

демо-версия вышла в 2005 г., однако активно использовать его стали только в 2007 г. Этот комплекс программ состоит из Selenium-Core, Selenium-RC и SeleniumIDE.

Selenium-Core – это браузерный бот, написанный на JavaScript. Его цель – запуск тестов в настоящем браузере. Тесты имитируют работу пользователя в браузере, причем все недостатки конкретного браузера также включены в этот тест, так как для теста используется самый обычный браузер. Сами тесты пишутся в форме HTML таблиц. Выглядит это примерно так:

```
<table>
  <tr>
    <td>open</td>
    <td>/change_address_form.html</td>
    <td></td>
  </tr>
  <tr>
    <td>type</td>
    <td>address_field</td>
    <td>Betelgeuse state prison</td>
  </tr>
  <tr>
    <td>clickAndWait</td>
    <td>//input[@name='Submit']</td>
    <td></td>
  </tr>
  <tr>
    <td>verifyTextPresent</td>
    <td>Address change successful</td>
    <td></td>
  </tr>
</table>
```

Как видно, формат теста довольно прост. Это таблица, в которой строки – последовательность команд, а столбцы – имя и параметры команды. Такие тесты можно собирать в наборы и запускать все или по отдельности. В дальнейшем эти тесты можно использовать для регрессионного тестирования при внесении изменений.

Такие тесты должны писаться на основе userstory и обязательно не тем человеком, который писал приложение.

Модульное тестирование

Модульное тестирование – процесс в программировании, позволяющий проверить на корректность отдельные модули исходного кода программы. Идея состоит в том, чтобы писать тесты для каждой нетривиальной функции или метода. Это позволит достаточно быстро проверить, не привело ли очередное изменение кода к регрессии, т.е. к появлению ошибок в уже написанных и оттестированных местах программы, а также облегчает локализацию и устранение таких ошибок. Цель модульного тестирования – изолировать отдельные части программы и показать, что по отдельности эти части работоспособны. Этот тип тестирования обычно выполняется программистами. Модульное тестирование обеспечивает следующие преимущества.

- Поощрение изменений. Модульное тестирование позже позволяет программистам проводить рефакторинг, будучи уверенными, что модуль по-прежнему работает корректно (регрессионное тестирование). Это поощряет программистов к изменениям кода, поскольку достаточно легко проверить, что код работает и после изменений.

- Упрощение интеграции. Модульное тестирование помогает устранить сомнения по поводу отдельных модулей и может быть использовано для подхода к тестированию «снизу вверх» – сначала тестируются отдельные части программы, затем программа в целом.
- Документирование кода. Модульные тесты можно рассматривать как «живой документ» для тестируемого класса. Клиенты, которые не знают, как использовать данный класс, могут использовать модульный тест в качестве примера.

Фреймворки

На сегодняшний день все фреймворки для модульного тестирования являются PHP-клонами фреймворка JUnit. Самые используемые – это PHPUnit и SimpleTest. Оба этих фреймворка реализуют простую концепцию. Есть класс – набор тестов. Все методы этого класса, название которых начинается с ключевого слова test, являются отдельными тестами. В тестах пишутся утверждения. Если утверждение оказывается ложным, то, значит, тест не успешен. Как сказано выше, эти наборы инструментов идентичны по идеологии, однако, они различаются по полноте реализации различных аспектов тестирования. Для удобства выбора между ними составлена таблица [2, 3].

Описание	PHPUnit	SimpleTest
Суперкласс для тестового набора и предположения	+	+
Результаты в HTML	+	+
Автоматическая загрузка тестовых наборов	-	+
mock-объекты	+	+
Набор тестов для веб	+	+
Частичные mock'и		+
Поддержка cookie	+	+
Поддержка редиректов	+	+
Парсинг форм	+	+
Запуск из командной строки	+	+
Вывод в XML	+	+
HTTP авторизация	+	+
Поддержка SSL	+	+
Поддержка прокси	+	+
Поддержка фреймов		+
Поддержка тестирования загрузки файлов	+	+
Тестирование исключений	+	+
Поддержка тега base	+	+
Покрытие кода	+	-
Взаимодействие с Selenium-RC	+	-
Тестирование производительности	+	-
Незаконченные тесты	+	-

Таблица. Сводные характеристики фреймворков для модульного тестирования

Как видно из таблицы, написание приемочных тестов для Selenium-RC доступно только в PHPUnit. Однако PHPUnit поставляется в виде PEAR модуля, поэтому установить его не на свой сервер невозможно. SimpleTest же целиком написан на PHP, и его можно самостоятельно поставить на хостинг. В версии SimpleTest для разработчиков тоже есть некоторая поддержка Selenium. Для примера напишем какой-нибудь тест.

Пусть необходимо написать шлюз для платежной системы, через которую пользователи оплачивают услуги компании. Платежная система шлет запросы шлюзу в виде GET-запросов. В ответ шлюз должен выдавать XML, подтверждающий получение платежа или же описывающий ошибку, которая произошла в процессе обработки запроса. Тестирующий модуль будет слать запросы шлюзу, имитируя тем самым платежную систему, считывать ответы и сравнивать с ожидаемыми.

```
<?php
define('SIMPLE_TEST', '../simpletest/');
require_once(SIMPLE_TEST . 'web_tester.php');
require_once(SIMPLE_TEST . 'reporter.php');

class PaymentTest extends WebTestCase {
function testValidCheck() {
    $answer_params = array( '{code}' => '0',
' {message}' => 'Абонент существует, возможен прием платежей'
);
    $this->get( $this->path .
'?action=check&number=5011&type=0' );
    $this->assertSource( $this->getAnswer(
$answer_params ) );
}
}
$paytest = new PaymentTest();
$paytest->run( new HtmlReporter( 'windows-1251' ) );
?>
```

Для примера мы взяли один тестирующий метод. Он тестирует успешную проверку возможности зачисления платежа и состоит из двух вызовов: отправка запроса к шлюзу и сравнение результата с эталоном. В случае несовпадения тест выдаст ошибку.

Непрерывная интеграция

Непрерывная интеграция заключается в регулярной проверке работоспособности приложения с помощью запуска автоматических тестов. Как правило, есть некоторый репозиторий, в который разработчики постоянно вносят результаты своей работы. При изменении либо при выполнении другого условия выполняются тесты. В случае провала теста ошибка находится и исправляется как можно быстрее. Таким образом, поддерживается работоспособность проекта в репозитории.

К сожалению, готовых решений для PHP найти не удалось.

Заключение

В статье рассматривались способы поддержания качества web-приложений, написанных на языке PHP. В настоящее время разработка ответственных проектов на PHP – уже не редкость. Таким образом, область применения предьявляет свои

требования к надежности таких систем. Для PHP имеется ряд средств, позволяющих поддерживать качество web-приложений, написанных на этом языке. Они включают в себя модульные тесты, web-тесты, тесты интерфейсов.

Литература

1. Кириллов А.В. Пишем PHP код, устойчивый к ошибкам. – Режим доступа: <http://www.compdoc.ru/internet/php/code/>, своб.
2. Overview of SimpleTest. – Режим доступа: <http://simpletest.sourceforge.net/en/overview.html>, своб.
3. Bergmann S. PHPUnit Manual. – Режим доступа: http://www.phpunit.de/pocket_guide/3.2/en/index.html, своб.

ОБЗОР РЕШЕНИЙ ДЛЯ УПРАВЛЕНИЯ ИТ-ИНФРАСТРУКТУРОЙ

Д.В. Пудов

Научный руководитель – к.т.н., доцент Б.А. Крылов

Представлен обзор решений в области управления ИТ-услугами. Проведен анализ масштабируемости и архитектуры, соответствия мировым стандартам в области управления ИТ-услугами.

Введение

Важность информационных технологий для бизнеса стремительно возрастает. Многие компании уже пришли к пониманию важности роли информационных технологий для их основного бизнеса. Сегодня невозможно предоставить пользователю определенные категории услуг, не поддерживая их в значительной степени информационными технологиями. Скорость изменений в бизнесе в целом стремительно возрастает. Происходит слияние компаний, поглощение, реструктуризация и т.п. Эти процессы требуют поддержки, ресурсов внутри компании, внутри основного бизнеса – того, за счет чего можно добиться определенных стратегических преимуществ на рынке. Сегодня именно информационные технологии как крупный сегмент, требующий значительных инвестиций, внимательно рассматривается в качестве ресурса, из которого могут быть извлечены определенные преимущества для бизнеса.

В этих условиях задача эффективного управления самой информационной инфраструктурой выходит на передний план. Только тогда, когда мы поймем влияние ИТ-операций на бизнес и приведем информационную инфраструктуру компании в соответствие с бизнес-процессами, можно достичь реальной отдачи от своих ИТ-ресурсов. Многие компании обнаружили, что самой большой трудностью, но и пользой во внедрении информационных технологий является объединение ключевых процессов для повышения отдачи. Использование информационных технологий повышает эффективность каждого процесса за счет удовлетворения его непосредственных потребностей.

Библиотека ИТИЛ

Общий подход к управлению информационными технологиями значительно эволюционировал в течение последних лет. Так, если раньше ИТ-организации фокусировались в основном на управлении фактическими системами и ресурсами, то с течением времени организации значительно сместили акцент на управление ИТ-услугами.

Процессы управления услугами описываются в двух разделах библиотеки ИТИЛ (англ. Information Technology Infrastructure Library, библиотека для консультантов, содержащая полученные на основе обобщения мирового опыта рекомендации и решения для создания инфраструктур и организации работы корпоративных систем): сервисная поддержка и предоставление услуг. В разделе сервисной поддержки определены методы обеспечения доставки услуг путем поддержки пользователей и контроля в рамках инфраструктуры над изменениями и релизами. В разделе предоставления услуг описывается доставка ИТ-услуг ИТ-заказчикам в соответствии с оговоренным уровнем обслуживания.

Процессы сервисной поддержки включают управление:

- **Инцидентами.** Управление инцидентами заключается в оперативной обработке инцидентов, зарегистрированных диспетчерской службой – включая неисправности, простои и запросы на обслуживание. Автоматизация процесса управления инцидентами включает регистрацию и отслеживание инцидентов, мониторинг состояния системных ресурсов, обработку стандартных запросов на обслуживание, а также управление базами знаний.

- **Проблемами.** В отличие от процесса управления инцидентами, ориентированного на решение краткосрочных задач, управление проблемами заключается в анализе накопленных данных с целью определения тенденций, выполнения анализа перво-причин и выявления долгосрочных решений ошибок организации ИТ-инфраструктуры. Поддержка автоматизации управления проблемами включает анализ накопленных данных, регистрацию проблем, их отслеживание и диагностику.
- **Изменениями.** В задачи процесса управления изменениями входят проверка и утверждение запросов на изменение, а также координация и развертывание утвержденных изменений.
- **Релизами.** Управление релизами – это процесс подготовки, тестирования и развертывания инфраструктурных изменений, называемых «релизами».
- **Конфигурациями.** Управление конфигурациями позволяет контролировать информацию об аппаратном и программном обеспечении и сопутствующей документации ИТ-инфраструктуры (называемых «учетными элементами»), а также о взаимосвязях между этими учетными элементами.

Процессы предоставления услуг включают управление:

- **Доступностью.** Управление доступностью подразумевает выяснение требований заказчика к уровню доступности услуг, оценку мощности ИТ-инфраструктуры для обеспечения этих уровней доступности и выполнение определенных действий для повышения уровня доступности услуг.
- **Мощностью.** Управление мощностями – это проактивная проверка наличия у ИТ-инфраструктуры мощностей, достаточных для эффективной работы.
- **Уровнем обслуживания.** Управление уровнями обслуживания заключается в ведении и отслеживании соглашений об уровнях обслуживания (SLA).
- **Непрерывностью предоставления ИТ-услуг.** Управление непрерывностью предоставления ИТ-услуг – это процесс планирования восстановления системы после основных сбоев ИТ-услуг.
- **Финансами в сфере ИТ.** Управление финансами в сфере ИТ охватывает финансовое управление вокруг управления ИТ-активами и ресурсами.

Обзор решений

Двумя основными лидерами рынка решений в области управления ИТ-инфраструктурой являются IBM Tivoli и HP Software. Рассмотрим эти решения с точки зрения следующих вопросов: возможности и соответствие ITIL, масштабируемость и архитектура, поддерживаемые платформы.

Библиотека ITIL является стандартом де-факто в области ITSM (англ. IT Service Management, управление ИТ-услугами), в связи, с чем ведущие производители стараются в полной мере соответствовать этому стандарту. Программные решения от IBM и HP наиболее полным образом реализует в себе поддержку ITIL.

Первоначально Tivoli имела традиционную двухуровневую архитектуру, состоящую из управляющего и управляемых объектов. На сегодняшний день ПО Tivoli реализует более прогрессивную трехуровневую архитектуру, когда между управляющим сервером и управляемым агентом (конечной точкой, endpoint) помещается управляющий шлюз (gateway), который обеспечивает связь между определенной группой управляемых агентов и остальной средой Tivoli. Управляющий шлюз является управляемым объектом с точки зрения управляющего сервера, однако отвечает за все коммуникации с подчиненными ему конечными точками без дополнительных соединений с управляющим сервером.

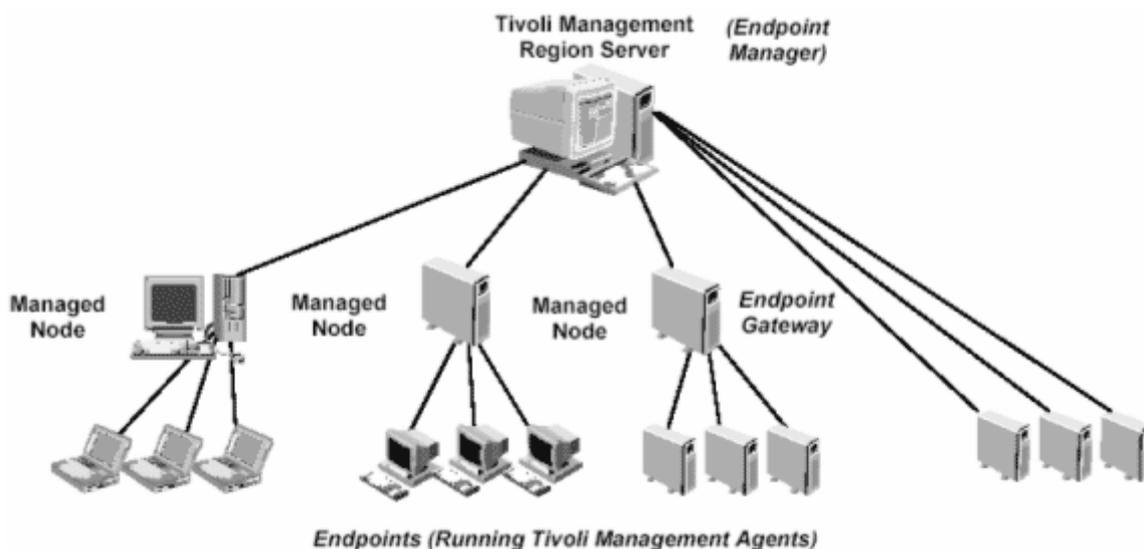


Рис. 1. Трехуровневая архитектура Tivoli

Трехуровневая архитектура Tivoli включает:

- Tivoli Management Region Server (TMR Server) – основной сервер системы управления. В его задачи входит обеспечение работы системы управления, поддержка внутренней базы данных объектов управления. TMR-Server – это первый компьютер, на который устанавливается Framework. Все объекты (ресурсы), которыми управляет TMR Server, считаются принадлежащими одному региону Tivoli. Два или более таких региона могут обмениваться информацией.
- Managed Node (MN) – сервер, позволяющий перераспределить нагрузку с TMR-сервера, что обеспечивает масштабируемость системы управления. TMR-Server поддерживает сотни MN, а любой MN (в том числе и TMR-Server) может использоваться в системе управления в качестве Endpoint Gateway – шлюза, через который осуществляется управление тысячами пользовательских компьютеров и серверов.
- Tivoli Management Agent (TMA) обеспечивает управление пользовательским компьютером или сервером. Агент в системе управления Tivoli Enterprise – единый, т.е., если в какой-то момент от системы управления потребуется дополнительная функциональность, то достаточно установить нужный модуль на TMR-Server, и агент получит необходимый код автоматически. Программный код агента использует минимум ресурсов системы и не мешает работе пользовательских приложений. Это единственный компонент, который необходимо установить на управляемую станцию или сервер.

Данная архитектура обеспечивает работу системы управления Tivoli в информационных системах практически любого масштаба. Сотни тысяч клиентских мест и серверов могут функционировать под управлением Tivoli, подчиняясь единым правилам. При этом с точки зрения системы управления и, главное, с точки зрения использования человеческих ресурсов управляющая операция будет выглядеть абсолютно одинаково вне зависимости от того, над какими операционными системами и над каким количеством серверов и рабочих станций она выполняется.

Архитектура HP Software реализует классическую двухуровневую архитектуру, но это не является помехой для управления огромными сетями. Например, для приложения HP Enterprise Discovery 2.51, управляющая часть которого может быть установлена только под Windows 2003 Server, предъявляются требования к аппаратному обеспечению, представленные на рис. 2.

Discovered Devices	Inventoried Devices	Ports	Aggregation	Automated Inventory	Topology	Memory (GB)	CPU*	Disk (GB)
6,000	5,000	36,000		x		1.5	1 CPU 2.4 GHz	25
6,000	5,000	36,000		x	x	3**	1 CPU 2.8 GHz	40
18,000	15,000	108,000		x		2	2 CPUs or cores 3.0 GHz	70
18,000	15,000	108,000		x	x	5	2 CPUs or cores 3.0 GHz	105
60,000	50,000	150,000		x		4	2+ CPUs or cores 3.6 GHz	200
60,000	50,000	150,000		x	x	7	2+ CPUs or cores 3.6 GHz	260
50,000	50,000	n/a	x	n/a	n/a	2	2 CPUs or cores 3.0 GHz	10
500,000	500,000	n/a	x	n/a	n/a	3	2 CPUs or cores	50

Рис. 2. Рекомендуемая аппаратная конфигурация для Enterprise Discovery

Заключение

Сегодня на мировом рынке решений ITSM имеется четыре основных лидера – IBM и HP, а также BMC и CA, но наиболее агрессивно играют IBM и HP. Эти компании инвестируют миллиарды долларов в разработку программного обеспечения, покупают более мелких игроков рынка. IBM приступила к агрессивным действиям на этом рынке в 1993 г., на 2 года раньше HP. Линейка продуктов каждого из игроков состоит из более чем 100 продуктов. Многие из них выполняют одинаковые функции и отличаются лишь некоторыми особенностями, которые могут иметь важное значение для одного предприятия и быть лишней для другого. Таким образом, определить объективного победителя нет возможности.

Литература

1. ITIL в контексте стандартизации [Электронный ресурс] / К. Поляков. – [Россия]: Изд-во «Открытые системы» – (журнал «Директор информационной службы»; 26/04/2007 №04) – Режим доступа: <http://www.osp.ru/cio/2007/04/4151978/>
2. IBM Tivoli Software library for technical resources [Electronic resource] / IBM Corp. – Режим доступа: <http://www-306.ibm.com/software/tivoli/sw-library/>
3. [Решения CompuTel, построенные на ПО IBM Tivoli] [Электронный ресурс] / системные интегратор CompuTel – Режим доступа: <http://www.tivoli.computel.ru/solutions/>
4. Решения для оптимизации бизнес-технологий (ВТО) [Электронный ресурс] / HP Corp. – Режим доступа: <http://www.hp.ru/openview/library/>

РАЗРАБОТКА ВЕБ-ПРИЛОЖЕНИЙ НА ОСНОВЕ ТЕХНОЛОГИИ AJAX В СРЕДЕ MICROSOFT VISUAL STUDIO

В.Н. Зимин, П.В. Федосов

Научный руководитель – к.т.н., доцент Б.А. Крылов

Технология AJAX позволяет реализовывать качественно новый тип веб-приложений. Это касается как увеличения скорости реакции на действия пользователя, так и сокращения объема данных, передаваемых между клиентской и серверной частями приложения. Применение специализированных сред разработки, например, Microsoft Visual Studio, значительно облегчает задачу разработки приложений такого типа.

Введение

В последнее время в разработке веб-приложений наметилась тенденция усложнения и увеличения объема кода, выполняемого на стороне клиента. Это вызвано как существенным усложнением задач, решаемых веб-приложениями, и расширением круга этих задач, так и изменением архитектурной концепции построения веб-приложений. На современном этапе все большая роль в работе веб-приложения возлагается на его клиентскую часть, что заставляет разработчиков делать его все более объемным, надежным и гибким. Все это, в свою очередь, потребовало кардинального изменения подхода к самому процессу разработки таких веб-приложений и появления новых инструментов и сред разработки, отвечающих растущим требованиям со стороны пользователей. Среди технологий, используемых в веб-приложениях, все большую популярность приобретает технология AJAX (Asynchronous Javascript and XML). Технология AJAX позволяет реализовывать качественно новый тип веб-приложений. Это касается как увеличения скорости реакции на действия пользователя, так и сокращения объема данных, передаваемых между клиентской и серверной частями приложения, что, в свою очередь, увеличивает скорость работы приложения и сокращает нагрузку на сервер. Технически разработка AJAX-приложений возможна и с использованием старых средств разработки, однако применение специализированных сред, одними из которых являются Microsoft Visual Studio последних версий, серьезно облегчает эту задачу. В данной работе рассматриваются некоторые особенности разработки AJAX-приложений в интегрированной среде разработки Microsoft Visual Studio.

Свойства интегрированных сред разработки

Интегрированная среда разработки представляет собой приложение, реализующее инструменты для разработки приложений для той или иной платформы, собранные в единый набор. Как правило, интегрированная среда разработки содержит такие типичные элементы, как редактор исходного кода (обычно поддерживающий подсветку синтаксиса и автоматическое форматирование, характерное для используемого языка программирования), компилятор или интерпретатор, дополнительные утилиты и отладчик. Кроме того, часто в состав интегрированных сред разработки входят дополнительные средства, такие, как, например, утилита контроля версий. Большинство современных сред ориентированы на работу с графическими операционными системами и сами используют графический интерфейс пользователя. Для облегчения создания интерфейса пользователя графических приложений современные среды разработки содержат средства визуального проектирования, в значительной степени автоматизирующие этот процесс. Так как современные языки программирования, как правило, являются объектно-ориентированными, для автоматизации и стандартизации работы с классами и объектами среды разработки содержатся специализированные инструменты, такие как инспектор объектов (object inspector) или обозреватель классов (class browser) [1].

Таким образом, можно прийти к выводу, что интегрированные среды разработки созданы для повышения производительности труда программистов. В значительной степени это достигается путем предоставления программисту возможности использования больших фрагментов готового кода, поставляемого в рамках компонентов. Типичная среда ориентированна на конкретный язык программирования и оптимизирована для максимальной поддержки концепций и парадигм, свойственных этому языку. Тем не менее, существуют и многоязыковые среды, позволяющие использовать в рамках единой среды преимущества того или иного языка в зависимости от задачи, стоящей перед разработчиком. Примером таких сред могут служить такие продукты, как NetBeans и Microsoft Visual Studio.

Потребность в создании интегрированных сред разработки возникла тогда, когда производительность труда программиста при использовании классических (не интегрированных) средств перестала отвечать растущим запросам индустрии. Первым языком, ориентированным на использование в рамках единой среды разработки, был один из вариантов Бейсика. Разумеется, эта среда не имела графического интерфейса пользователя и других типичных для сегодняшних продуктов этого класса элементов. Однако она объединяла в себе средства редактирования исходного программного кода, управления файлами, компиляции, выполнения и отладки программ, т.е. имела все основные черты, характерные для современных приложений такого класса.

Стоит так же отметить, что, несмотря на очевидные преимущества, даваемые интегрированными средствами разработки в сравнении с классическими средствами, их внедрение не сразу было воспринято положительно использовавшими их разработчиками программного обеспечения. Причина этого кроется в том, что из-за стандартизации подходов к разработке программных продуктов, лежащих в основе самой концепции интегрированной среды, поле для творчества разработчиков неизбежно сужалось. Это стало одним из неизбежных следствий внедрения новой модели разработки программного обеспечения [1].

На современном этапе все большее распространения получают системы визуального программирования. Особенность визуальных интегрированных сред заключается в том, что процесс программирования в значительной степени автоматизируется за счет использования «строительных блоков» из готового кода, которые используются для сборки приложений. Причем это касается не только визуального проектирования интерфейса пользователя. Собранные воедино компоненты, графически представляемые в процессе разработки в виде диаграмм или иных образов, затем обрабатываются компилятором аналогично классическому исходному тексту.

Используя гибкость и мощь такого подхода в области распределенных систем и широкие возможности современных веб-браузеров (таких как Internet Explorer и Firefox), можно с минимальными трудозатратами создавать достаточно мощные веб-приложения. Такой подход используется также в специализированных программах, таких, как Openlab, где конечным пользователям требуется гибкость и широкие возможности полноценного языка программирования, но без проблем, связанных с обучением ему [1].

Среда Microsoft Visual Studio 2008

Одной из самых современных интегральных сред разработки на сегодняшний день является продукт Microsoft Visual Studio 2008. Среда разработки Microsoft Visual Studio – это набор инструментов и средств, предназначенных для помощи разработчикам программ любого уровня квалификации в решении сложных задач и создания новаторских решений. Роль Visual Studio заключается в том, чтобы улучшить процесс разработки и упростить разработку высокоэффективных программ. Следует отметить

высокопроизводительные редакторы кода, поддержку технологии IntelliSense, мастеров и различных языков кодирования в одной интегрированной среде разработки. Разработчики, применяющие Visual Studio, получают в свое распоряжение интегрированный продукт, включающий инструменты, серверы и службы. В Visual Studio содержатся инструменты для всех этапов разработки программного обеспечения (разработка, тестирование, развертывание, интеграция и управление). Visual Studio 2008 в сравнении с предыдущими версиями имеет переработанный пользовательский интерфейс. Общий вид главного окна Microsoft Visual Studio 2008 представлен на рис. 1 [2].

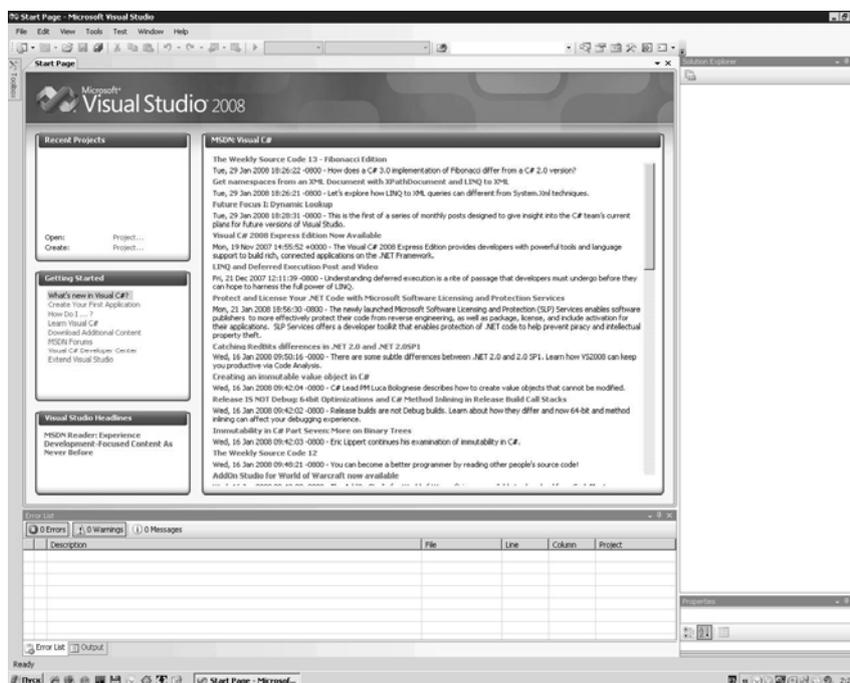


Рис. 1. Вид главного окна VS 2008

Технология AJAX

Суть технологии AJAX (Asynchronous Javascript and XML) заключается в изменении содержимого загруженной веб-страницы без ее полной перезагрузки, благодаря чему достигается высокая динамичность сайтов. Технология основывается на разделении данных и подзагрузки тех или иных компонентов по мере необходимости [3].

С архитектурной точки зрения технология AJAX представляет собой совокупность двух элементов: библиотеку клиентских скриптов и набор сервисных расширений. Библиотека клиентского кода реализована на языке JavaScript и, таким образом, совместима со всеми современными браузерами. Серверные расширения полностью интегрированы с серверными сервисами, элементами управления и т.д. [4].

С другой стороны, AJAX – это не самостоятельная технология, а, скорее, концепция использования нескольких смежных технологий. AJAX базируется на двух основных принципах:

- использование технологии динамического обращения к серверу «на лету», без перезагрузки всей страницы полностью (например: с использованием XMLHttpRequest; через динамическое создание дочерних фреймов; через динамическое создание тега <script>);
- использование DHTML для динамического изменения содержания страницы.

В качестве формата передачи данных обычно используются JSON или XML. Впервые термин AJAX был публично использован в 2005 г., однако в той или иной

форме многие технологии были доступны и использовались гораздо раньше, например, в подходе «Remote Scripting», предложенным компанией Microsoft в 1998 г., или с использованием HTML элемента IFRAME, появившегося в Internet Explorer 3 в 1996 г.

Технология AJAX имеет следующие основные преимущества.

- Экономия трафика. Использование AJAX позволяет значительно сократить трафик при работе с веб-приложением благодаря тому, что часто вместо загрузки всей страницы достаточно загрузить только небольшую изменившуюся часть.
- Уменьшение нагрузки на сервер. AJAX позволяет несколько снизить нагрузку на сервер. К примеру, в Gmail, при работе с письмами, серверу достаточно внести изменения в базу данных и отправить клиентскому скрипту сообщение об успешном выполнении операции без необходимости повторно создавать страницу и отсылать ее клиенту.
- Ускорение реакции интерфейса. Поскольку нужно загрузить только изменившуюся часть, то пользователь видит результат своих действий быстрее.

Недостатки AJAX:

- Интеграция со стандартными инструментами браузера. Динамически создаваемые страницы не регистрируются браузером в истории посещения страниц, поэтому не работает кнопка «Назад», предоставляющая пользователям возможность вернуться к просмотренным ранее страницам. Другой недостаток изменения контента страницы при постоянном URL заключается в невозможности сохранения закладки на желаемый материал.
- Динамически загружаемое содержание недоступно поисковикам. Поисковые машины не могут выполнять JavaScript, поэтому разработчики должны позаботиться об альтернативных способах доступа к содержимому сайта.
- Старые методы учета статистики сайтов становятся неактуальными. Многие сервисы статистики ведут учет просмотров новых страниц сайта, для сайтов, страницы которых широко используют AJAX, такая статистика теряет актуальность [5].

Разработка AJAX-приложения в Microsoft Visual Studio

Итак, обычно, когда программист начинает писать веб-приложение на основе AJAX предполагается, что у него уже есть как минимум базовое представление об объекте XMLHttpRequest и о том, как его создать и использовать. Но одно дело, когда речь идет об отправке пары HTML-форм, которые несложно реализовать вручную, и совсем другое, когда создается большое приложение, где только форм сотни.

Microsoft Visual Studio 2008 предоставляет уникальные возможности для создания приложений подобного рода. Благодаря наличию большого числа встроенных компонентов веб-разработчик избавлен от необходимости самостоятельно реализовывать многие компоненты, такие как календарь, часы, а порой и целые формы. На рис. 2 показано рабочее окно Visual Studio с подключенным набором Ajax элементов управления, разработанных фирмой Telerik [6].

Данный набор дополнений реализует свыше 400 различных элементов управления, которые позволяют прозрачно добавлять интерактивность фактически в любой проект. Также немаловажным преимуществом использования Visual Studio является то, что любое обычное веб-приложение легко хотя бы частично перевести на технологию AJAX в максимально сжатые сроки.

Немаловажную роль играет глубокая интеграция Visual Studio с другими компонентами, например Internet Information Services – сервером приложений, позволяющим проводить отладку веб-приложений. А тесная интеграция Visual Studio с Bug Tracking системами и отлаженная методика проведения Regression Testing позволяет всегда быть уверенным в качестве получаемого результата.

ПРЕДСТАВЛЕНИЕ ИНФОРМАЦИОННЫХ ПОТОКОВ С ПОМОЩЬЮ ТЕОРИИ ГРАФОВ

Т.С. Николаева

Научный руководитель – к.т.н., доцент Н.С. Кармановский

Правильная организация конфиденциального делопроизводства на крупных предприятиях является составной частью комплексного обеспечения безопасности информации и имеет огромное значение в достижении цели ее защиты. В статье рассмотрена модель информационных потоков, циркулирующих на предприятии, на основе теории графов. Использование такой модели позволяет своевременно выявить «проблемные» участки, затрудняющие сбор, обработку, использование и передачу конфиденциальной информации, которые, в свою очередь, являются неотъемлемыми элементами успешной деятельности любой организации.

Введение

Переход России к устойчивому, безопасному, конкурентоспособному экономическому росту возможен путем рационального управления информационными ресурсами как способом обеспечения эффективной деятельности предприятия. Создание информационных механизмов становится особенно актуальным, когда информация и знания трактуются как важнейший и редкий ресурс, как элемент экономического потенциала хозяйственной системы, который должен эффективно использоваться для достижения ее конкретных целей.

В практической деятельности часто приходится решать ряд вопросов, по существу, одной и той же задачи, когда варианты определяются не изменением значений исходных параметров, а спецификой решения в различных ситуациях. Такая специфика в решении задач особенно характерна для задач документооборота.

Для однотипных задач, не имеющих существенных различий, целесообразно использовать единый составной процесс. Такой процесс может отражать специфику решения задач в различных подразделениях, но быть настолько общим, чтобы обеспечить общность реализации. Следует обратить внимание на тот факт, что составной процесс является не простым соединением отдельных процессов, а результатом их логического объединения. Такое объединение выполняется с использованием специального математического аппарата.

Графы являются существенным элементом математических моделей в самых разнообразных областях науки и практики. Они помогают наглядно представить взаимоотношения между объектами и событиями в сложных системах [1]. Применение графов позволяет использовать апробированный аппарат для представления потоков данных и изменений состояний документооборота.

Сегодня, после экспоненциального увеличения возможностей компьютерной техники и технологий программирования, появилась возможность создавать более насыщенные и вместительные эффективные распределенные системы. С учетом открывшихся перспектив в настоящей статье будет рассмотрена модель построения информационных потоков с помощью теории графов.

Задание базиса

Для построения графовой модели необходимо определить набор данных, которые будут приняты за базис. Для графовой модели данные должны быть дискретны и предполагать связность различной степени [2]. За основу для рассматриваемой графовой модели возьмем определение информационных потоков. «Информационный поток – это множество действий, производимых множеством участников над множеством документов» [3]. В соответствии с данным определением информационные потоки можно

представить в виде: $I = \{P, U, D\}$, где I – модель информационных потоков, P – множество действий, U – множество участников, D – множество документов (рис. 1).

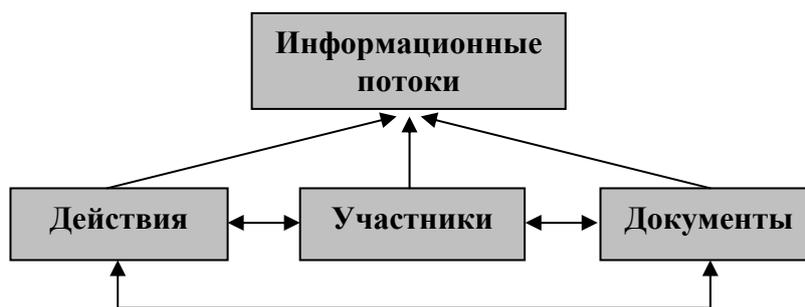


Рис. 1. Модель информационных потоков

При проведении анализа создаваемой модели информационных потоков необходимо учитывать связи между состояниями, т.е. последовательность действий, которые приводят систему к смене одного состояния на другое (преобразование документов от начальных состояний к требуемым (конечным)). Под определением действия будем понимать событие, возникновение которого приводит к изменению состояния хотя бы одного документа. Каждое состояние документа для своего возникновения требует некоего события, т.е. в системе должно произойти действие, которое приведет к возникновению этого состояния. Таким образом, чтобы провести множества P , необходима связность элементов из множества D .

Множество участников U предполагает также и множество ролей, которые они выполняют в системе документооборота. Современная управленческая модель документооборота подразумевает множественность ролей, выполняемых одним сотрудником. Поэтому под участниками следует понимать не конкретных персоналий, а перечень выполняемых ими должностных обязанностей. Все сотрудники в рабочем процессе выполняют несколько ролей, которые им назначены. В то же время одна и та же роль может соответствовать сразу нескольким сотрудникам. Таким образом, ролевое представление должностных обязанностей сотрудника дает возможность формализовать обязанности в виде множества ролей.

Множество D формируется путем выявления допустимых форм документов. Допустимые документы включают электронные версии бумажных документов и новые формы, возникновение которых связано с внутренними потребностями. Каждая из таких форм состоит из predetermined набора полей и начальных данных. Предполагается, что форма будет неизменной во время жизненного цикла документооборота. Это говорит о том, что если происходит изменение данных формы, то эта форма переходит в новое состояние.

Рассмотрим отношения между субъектами документооборота, которые определяют связи между элементами множеств. Отношения могут быть как между различными элементами одного множества, так и между элементами различных множеств. Определение ролей, т.е. влияние участников на конкретные состояния, приводящие к их изменению, определяется отношением элементов из множества U к элементам из множества D . В рамках нашей модели можно говорить, что отношения между элементами множеств задают отношения, совокупность которых определяет полный перечень процессов организации. Таким образом, мы исходим из того, что информационные потоки организации заданы в виде систем трех множеств, каждое из которых содержит конечное количество элементов.

Построение графовой модели

При построении графовой модели информационных потоков предлагается использовать следующий способ отображения. Для задания множества вершин графа будем использовать множество возможных состояний D . Ребра графа зададим с помощью множества действий P . Установим это соответствие таким образом, чтобы выполнялись следующие правила:

- одной вершине графа соответствует один и только один элемент множества D ;
- одному ребру графа соответствует один и только один элемент множества P ;
- одному элементу множества D соответствует одна и только одна вершина графа;
- одному элементу множества P соответствует одно и только одно ребро графа.

Такое тождественное отображение множеств состояний D в множество вершин v и множества состояний P в множество ребер e можно математически определить следующим образом: для любого i справедливо утверждение (1):

$$v(i) = D(i) \text{ и } e(i) = D(i), \quad (1)$$

где $i \in I, I = 1, 2, 3 \dots n$. Таким образом, определяется две парных грамматики – первая грамматика для установления перевода D в v , вторая грамматика для установления перевода P в e . Связи между вершинами тождественно соответствуют связям состояний моделируемого информационного потока. Вершины графа соединяют ребра в том и только в том случае, если соответствующие вершинам состояния связаны действием, соответствующим ребру, то есть $e = \{e, \text{ если ребро существует; } 0, \text{ если ребро отсутствует}\}$.

Направленность ребер устанавливается таким образом, чтобы отображать логику последовательности смены состояний информационных потоков. Вершина i является входящей вершиной для вершины j через ребро k в том и только в том случае, если состояние i сменяется на состояние j после совершения действия k . Таким образом, состояниям y_1, y_2, y_3 сопоставляются вершины графа $v_1, v_2 \dots v_n$, и каждая пара вершин v_i и v_j соединена дугой e_{ij} , идущей от v_i к v_j , в том и только том случае, когда состояние v_i является входным состоянием для v_j .

Типы графа

Для наглядного представления модели информационных потоков предлагается использовать два основных вида графов – ориентированные и неориентированные. В большинстве современных реализаций электронного документооборота используются только ориентированные графы, что накладывает ряд ограничений на применимость решения. В частности, на раннем этапе надо иметь четкое представление о направленности протекающих процессов, что на практике часто является очень сложным процессом. Рассмотрим целесообразность и адекватность применения различных видов графов в модели информационных потоков.

Неориентированные графы удобно использовать на этапах анализа и проектирования для наглядного отображения полученных при обследовании данных. Характерной для этих этапов особенностью является слабая связность и неустойчивость корреляций первичных данных. Модели начинают строиться на основании данных, полученных при первичном анализе. При выявлении дополнительных данных становятся явными корреляции, которые упраздняют предыдущие. В описанной ситуации неориентированный граф очень удобен для использования, так как позволяет лишь констатировать факт наличия связи между отношениями, не требуя установления направленности. Первые данные, полученные при анализе, вообще представляют собой множество состояний документа, что отображается вырожденным неориентированным графом. По мере поступления дополнительных данных становятся явными существующие отноше-

ния и начальные состояния, из которых начинаются бизнес-процессы. Это отображается слабосвязным неориентированным графом [2].

Ориентированные графы целесообразно использовать на этапах проектирования, реализации, внедрения и разработки. При разработке модели информационных потоков на вышеописанных этапах на неупорядоченные отношения между состояниями накладываются правила, описывающие их последовательность. При формализации этих правил важно обеспечить сохранение полученной информации о причинно-следственных связях. Эта информация наглядно и полно отображается с помощью ориентированных графов. При составлении графовых моделей бизнес-процессов удобно использовать циклы для отображения реальных процессов, происходящих на предприятии. На практике часто в производственном процессе используется циклическая организация, т.е. документ попадает в цикл, образованный между несколькими исполнителями и состояниями, который заканчивается по факту выполнения достаточных условий. Такая форма прозрачна и широко распространена в реальной жизни, тем не менее, она значительно усложняет задачу моделирования с точки зрения конечности моделируемых процессов. Не представляется возможным гарантировать факт возникновения условий достаточности, т.е. критериев окончания цикла. В таком случае всегда остается вероятность того, что цикл не будет завершён в пределах жизненного цикла документооборота. Таким образом, можно утверждать, что применение графов для описания модели информационных потоков накладывает неоправданные ограничения на синтезируемые системы и значительно сокращает функциональные возможности будущих систем [1].

Исходя из изложенного, можно утверждать, что для отображения процессов документооборота целесообразно использовать ориентированные графы, содержащие циклы.

Заключение

На основе методологии построения систем документооборота в статье представлена графовая модель информационных потоков, которая учитывает потоки движения документов на множество участников процесса и множество действий.

На основании описанной в статье модели информационных потоков возможно построение программного обеспечения, которое будет использовать приведенный выше аппарат теории графов для решения практических задач документооборота предприятий и организаций.

Литература

1. Алексеев В.Е., Таланов В.А. Графы и алгоритмы. Структуры данных. Модели вычислений: Учебник – М.: Интернет-Университет Информационных Технологий; Бином. Лаборатория знаний, 2006. – 320 с.
2. Ахо А.В., Хопкрофт Д.Э. и др. Структуры данных и алгоритмы – М.: Вильямс, 2007. – 400 с.
3. Некраха А.В., Шевцова Г.А. Организация конфиденциального делопроизводства и защита информации – М.: Академический проект, 2007. – 224 с.

РАЗРАБОТКА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ОБСЛУЖИВАНИЯ АБОНЕНТОВ СЕТИ IP-ТЕЛЕФОНИИ НА БАЗЕ ПРОТОКОЛА SIP

В.А. Козак

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

Произведен обзор современных систем обслуживания абонентов. Дана краткая характеристика IP УАТС Asterisk и протоколу SIP. Разработан сервер автоматизированного доступа к услугам и клиентские приложения к нему, представлен алгоритм обработки пользовательского вызова.

Введение

Системы обслуживания вызовов – call- и контакт-центры – как технология и специфический сегмент рынка появились в тот момент, когда у компаний возникла необходимость обработки большого объема обращений потребителей. Развитие компьютерных технологий позволило соединить операторские центры с базами данных и информационными системами. Такой подход особенно успешно реализуется в сфере бизнеса, где общение в значительной степени стандартизировано, где каждый контакт может быть отнесен к конкретному типу и зафиксирован для дальнейшего использования.

Услуги сетевого contact-центра обеспечивают:

- балансировку нагрузки на основе статистики реального времени и доступности операторов, состояния очередей вызовов, ограничений контролируемых параметров, а также даты (день недели, месяц и т.п.);
- идентификацию вызывающих абонентов contact-центра на уровне сети на основе доступа к БД, ANI, DNIS, что позволяет контекстуально подключать настраиваемые приветствия, обеспечивать выбор услуг и предпочтений, а также использовать самообслуживание в речевом портале;
- парковку звонка на уровне сети с подключением речевого портала, обеспечивая голосовое сопровождение во время ожидания, речевую навигацию с использованием алгоритмов распознавания и синтеза речи;
- интеллектуальную маршрутизацию вызовов на основе данных о звонке (ANI, DNIS), клиентского профиля звонящего (БД заказчика), навыков и профилей операторов contact-центра, выбора в интерактивном меню IVR, а также требуемого уровня обслуживания;
- интеграцию сетевой инфраструктуры заказчика услуг contact-центра с конвергентными сетями оператора [1].

Лидеры рынка систем обслуживания вызовов

Лидерами рынка call-центров на сегодняшний день являются четыре компании – Avaya, Nortel Networks, подразделение компании Alcatel – Genesys и Cisco Systems. Присутствие компаний на рынке представлено в табл. 1.

	Avaya	Genesys	Nortel	Mitel	Cisco Systems	Interactive Intelligence
Кол-во установленных центров	25000	3000	14000	2100	1700	1000
Число работающих операторов	6-8 млн.	1,25 млн.	4 млн.	500 тыс.	35 тыс.	100 тыс.

Таблица 1. Присутствие компаний на рынке Call-центров в 2005 году

Для современной системы характерна работа под управлением производительной серверной операционной системы (Win2003, Sun Solaris, IBM AIX), интеграция с мощной и надежной СУБД и CRM-системой. Необходима реализация различных мультимедийных задач и задач доступа в Web. В табл. 2 представлены характеристики систем ведущих вендоров.

	Avaya (Interaction Sever 7.0)	Nortel (Symposium Suite, rel. 5)	Genesys (Genesys 7 Suite)	Mitel (6100 Contact Center Solutions v4.5)
Рекомендуемые серверы	Sun Solaris, IBM AIX и/или Windows	Win2000, частично Win2003	Win2000 primarily, Sun Solaris, IBM AIX, Linux	Win2000, Win2003
СУБД	Oracle, IBM DB2 SQL	Sybase RDBMS	SQL-based RDBMS, MS SQL Server	MS SQL Server
Типичное количество операторов	100-200	300	500	50
Максимальное количество операторов	5200	3300	16000	300-350
Поддержка PBX	Avaya CommMgr (на Linux), Siemens, Nortel, Ericsson, Aspect ACD	Nortel PBX	Server v7.0 (Genesys), Cisco, Nortel, Avaya и более 20 других	Mitel PBX
Поддержка Email	есть	есть	есть	есть
Внутренняя связь	Внутренняя (агент-агент-супервизор)	Чат может быть использован для внутренних переговоров	Чат может быть использован для внутренних переговоров	Поддерживает
Web-чат	есть	есть	есть	есть
Web-поддержка	есть	есть	есть	есть
Интеграция с CRM	SAP, Siebel, PeopleSoft, Onyx и Epiphany	MS CRM, SAP, PeopleSoft и Siebel	Собственные продукты, MS CRM, SAP, PeopleSoft, Siebel	MS CRM, SAP, АСТ, PeopleSoft, Remedy, Heat
Цена ПО на одного оператора	\$1377 - \$2877	\$1270	\$1725 (только голосовой) до \$2800 – полный пакет	\$805 до \$2200 при полном пакете мультимедиа

Таблица 2. Характеристики ведущих call-центров

На рынке платформ для call-центров также активно присутствуют такие телекоммуникационные гиганты, как Ericsson с платформой Next Call Centre, Siemens с решением HiPath ProCenter Standard и другие.

Выбор телекоммуникационной платформы

Рассмотренные выше системы, при всех их достоинствах, достаточно дороги. А поскольку в мире Open Source существует ряд открытых программных коммутаторов (УАТС), то задача разработки на их основе новой системы обслуживания выглядит довольно таки актуально.

Решение мультимедиа задач в разрабатываемой системе отводится IP УАТС. Роль этого компонента в данной системе трудно переоценить. Он является носителем интеллектуальных возможностей телефонной сети, координирует управление обслуживанием вызова, сигнализацию и функции, обеспечивающие установление соединения в одной или нескольких сетях. В первую очередь, коммутатор управляет обслуживанием вызовов, то есть установлением и разрушением соединений. В число функций управления обслуживанием вызова также входят распознавание и обработка цифр номера для определения пункта назначения вызова; а также распознавание момента ответа вызываемой стороны, момента, когда один из абонентов кладет трубку, и регистрация этих действий для начисления платы [2]. IP УАТС координирует обмен сигнальными сообщениями между сетями, то есть поддерживает функции Signaling Gateway. Иначе говоря, это устройство координирует действия, обеспечивающие соединение с логическими объектами в разных сетях и преобразует информацию в сообщениях с тем, чтобы они были понятны на обеих сторонах несхожих сетей. Результаты обзора корпоративных АТС сведены в табл. 3.

IP PBX	Платформа	Основные протоколы сигнализации	Лицензия/Стоимость	Важные дополнительные возможности
Asterisk	Linux, BSD, OS X	ADSI, SIP, H.323, MGCP, SCCP/Skinny	двойная лицензия, в т. ч. GPL	Voicemail, IVR, Media Gateway, Call Queuing, факс
CallWeaver	Linux, BSD, OS X, Solaris	ADSI, SIP, H.323, MGCP, SCCP/Skinny	GPL	Voicemail, IVR, Media Gateway, Call Queuing поддержка STUN
sipX	Linux	SIP	LGPL	Voicemail, Media Gateway, Call Queuing
AlterCall Switch	Linux	H.323, SIP	1700 руб/канал	Media Gateway, Call Queuing
SGpbx	Linux	H.323, SIP	2000 руб/канал	Media Gateway, факс, QoS

Таблица 3. Сравнение программных IP УАТС

В качестве коммутатора предполагается использовать Asterisk. Asterisk – это полноценная программная АТС, Open Source проект компании Digium. Она может работать на таких операционных системах, как Linux, BSD и OS X и предоставляет все возможности, которые есть у обычной мини-АТС. Функционирование Asterisk основано на протоколах, которые обеспечивают передачу голоса через сети, основанных на IP протоколе (VOIP), и благодаря этому данная АТС может работать практически с любым оборудованием для IP-телефонии. Asterisk предоставляет функции голосовой почты (Voicemail), конференций, интерактивного голосового меню (IVR), центра обработки вызовов (Call Queuing), поддержку протокола T.38 (факс). Он также имеет поддержку таких сервисов, как перевод вызовов другому абоненту, сервис определения и пере-

дачи вызываемому абоненту номера вызывающего абонента, протоколы ADSI, SIP, H.323, MGCP (только для call manager) и SCCP/Skinny [3].

Проанализировав современные протоколы сигнализации, было решено использовать протокол инициирования сеансов (SIP – Session Initiation Protocol, RFC 3261). SIP является протоколом прикладного уровня и предназначается для организации, модификации и завершения сеансов связи: мультимедийных конференций, телефонных соединений и распределения мультимедийной информации. Пользователи могут принимать участие в существующих сеансах связи, приглашать других пользователей и быть приглашенными ими к новому сеансу связи [1]. На рис. 1 приведен пример возможного построения сети SIP.

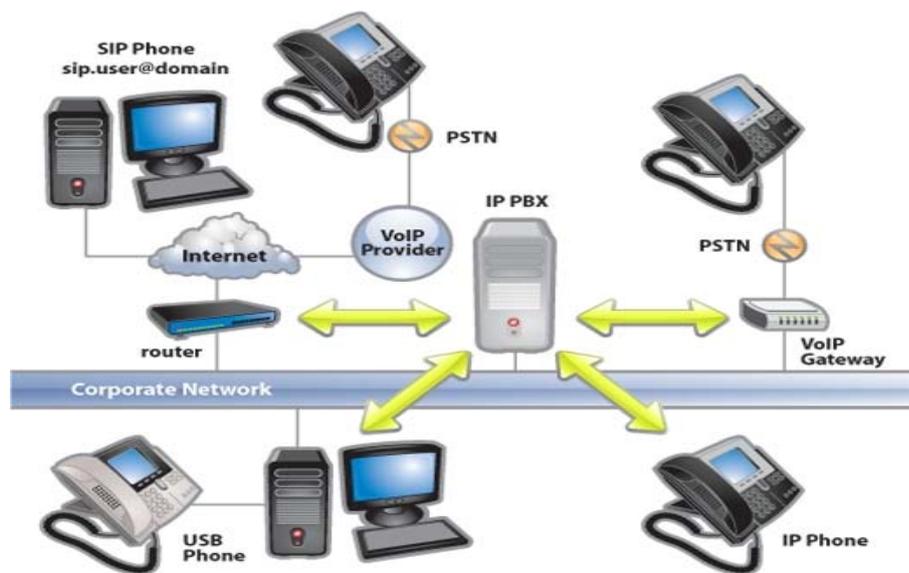


Рис. 1. Пример построения сети SIP

Разработка сервера автоматизированного доступа к услугам

В состав системы должны войти сервер автоматизированного доступа к услугам (данный компонент, выполняющий бизнес-функции системы, необходимо разработать), производительная и надежная IP УАТС, современная СУБД. Операторское приложение исполняется в виде тонкого клиента.

Перед сервером автоматизированного доступа к услугам ставятся задачи:

- управления коммутатором посредством протоколов телефонной сигнализации;
- взаимодействие с операторскими телефонами;
- взаимодействие с операторами по каналам web;
- выполнение бизнес задач по обслуживанию абонентов;
- управление информацией об абонентах, операторах, администраторах, услугах;

Разрабатываемый сервер представляет из себя J2EE-приложение. Java 2 Platform, Enterprise Edition (J2EE) определяет стандарт для разработки многоуровневых корпоративных приложений. J2EE упрощает разработку корпоративных приложений за счет использования стандартизованных модульных компонент и обеспечения полного набора сервисов для этих компонент. Многие аспекты работы приложения выполняются автоматически, не прибегая к сложному программированию. Данная технология поддерживается всеми ведущими производителями серверов приложений, что гарантирует переносимость и возможность интеграции с другими системами [4].

В состав сервера входят следующие логические блоки:

- блок управления;

- компонент управления вызовами;
- компоненты EJB;
- Web-компоненты.

Компонент управления – логический центр системы, он ответственен за принятие решений. Концентрируя в себе все остальные части системы, данный компонент обладает всей совокупностью информации о происходящих в системе процессах.

Компонент управления вызовами необходим для взаимодействия системы с другими SIP элементами. Если говорить терминами SIP, этот компонент выполняет функции User Agent и B2BUA. Компонент отвечает за формирование SIP-пакетов и отправку их к коммутатору и операторским телефонам. При поступлении входящего SIP-пакета происходит его обработка и уведомление блока управления о произошедшем в телефонной сети.

Компоненты EJB представляют из себя совокупность нескольких бинов сущности (Enterprise Bean), сессионных бинов (Session Bean) и бинов, управляющих сообщениями (Message Bean). *Бин сущности* представляет бизнес-объект в постоянной памяти. По сути, это способ надежного хранения информации. А сессионные бины – способ управления этой информацией. Бины, управляющих сообщениями, входят в систему асинхронных сообщений, позволяющую двум или более приложениям обмениваться информацией в форме сообщений [4].

Web-компоненты представлены как ставшими уже привычными страницами JSP и сервлетами, так и приложением, построенным с использованием технологии GWT (Google Web Toolkit). Данная технология позволяет писать AJAX-приложения на языке Java. AJAX (Asynchronous JavaScript and XML – «асинхронный JavaScript и XML») – это подход к построению интерактивных пользовательских интерфейсов веб-приложений, заключающийся в «фоновом» обмене данными браузера с веб-сервером. В результате при обновлении данных веб-страница не перезагружается полностью, и веб-приложения могут быть сделаны более быстрыми и удобными. После трансляции GWT-приложение превращается в совокупность файлов html, css и javascript, которые при пользовательском запросе отправляются по протоколу HTTP и исполняются браузером [5]. Все рассмотренные компоненты компилируются и собираются в единое Enterprise приложение.

Структура базы данных

Для хранения постоянной информации используется реляционная база данных. Схема базы данных показана на рис. 2.

В базе присутствуют таблицы:

- ORGANIZATION – таблица организации, обслуживаемых системой;
- EMPLOYEE – таблица сотрудников организации;
- SERVICE – таблица услуг, предоставляемых организацией;
- employebean_services_servicebean_employees – таблица для связи многие ко многим между EMPLOYEE и SERVICE;
- USERSESSIONCONTEXT – таблица с информацией о сотрудниках, использующих систему;
- WORKITEM – таблица задач на обслуживание;
- ENDUSER – таблица конечных пользователей (абонентов).

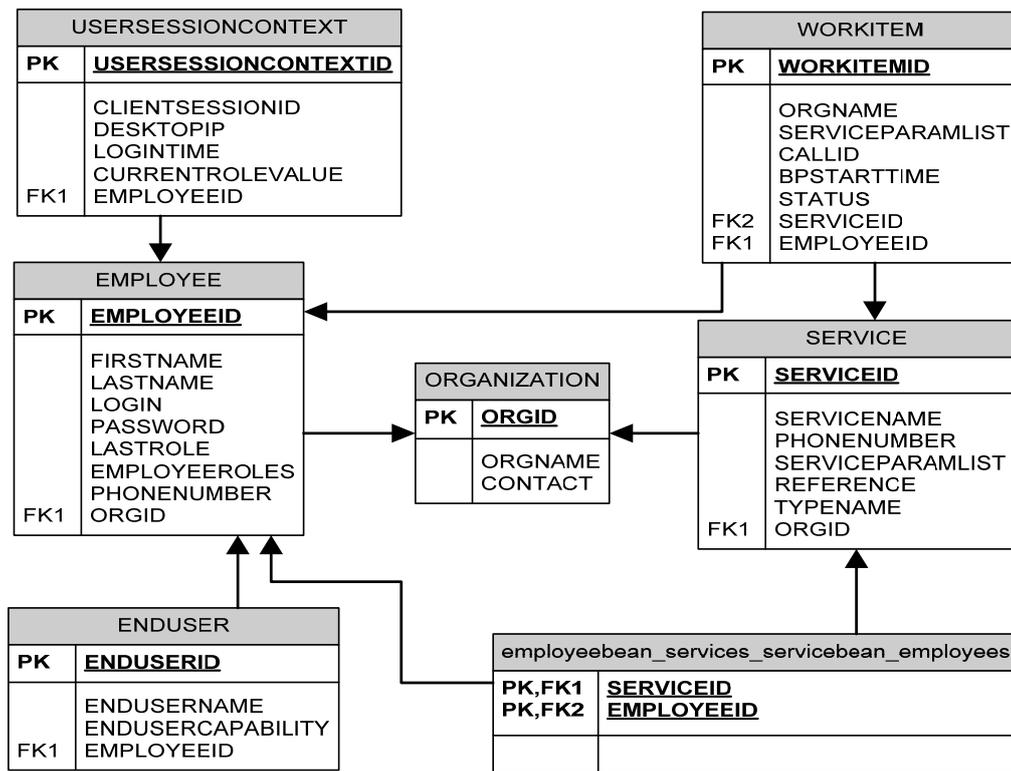


Рис. 2. Структура базы данных

Развертывание

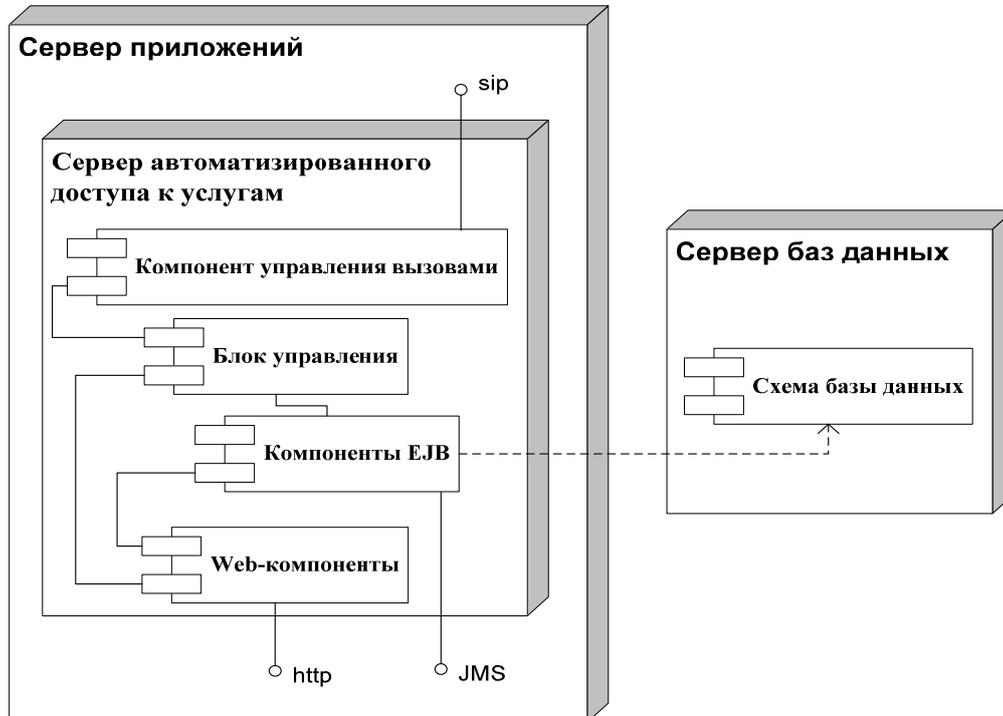


Рис. 3. Диаграмма развертывания

Разработанное приложение развертывается на двух логических серверах (желательно установленных на различных физических серверах):

1. сервер баз данных MySQL (использовалась версия 5.0.27);
2. сервер приложений JBoss (использовалась версия 4.2).

На сервер баз данных устанавливается схема базы данных. На сервер приложений устанавливаются «сервер автоматизированного доступа к услугам». Диаграмма развертывания системы приведена на рис. 3.

Алгоритм обработки пользовательского вызова

Главная задача создаваемой системы – обслуживание входящих вызовов. В алгоритме обслуживания вызова, приведенном на рис. 4, есть ряд ключевых моментов:

- по вызываемому номеру телефона определить услугу, в которой нуждается потребитель;
- по вызывающему номеру телефона определить, разрешено ли этому потребителю пользоваться данной услугой;
- определить операторов, способных принять данный вызов;
- принять вызов.

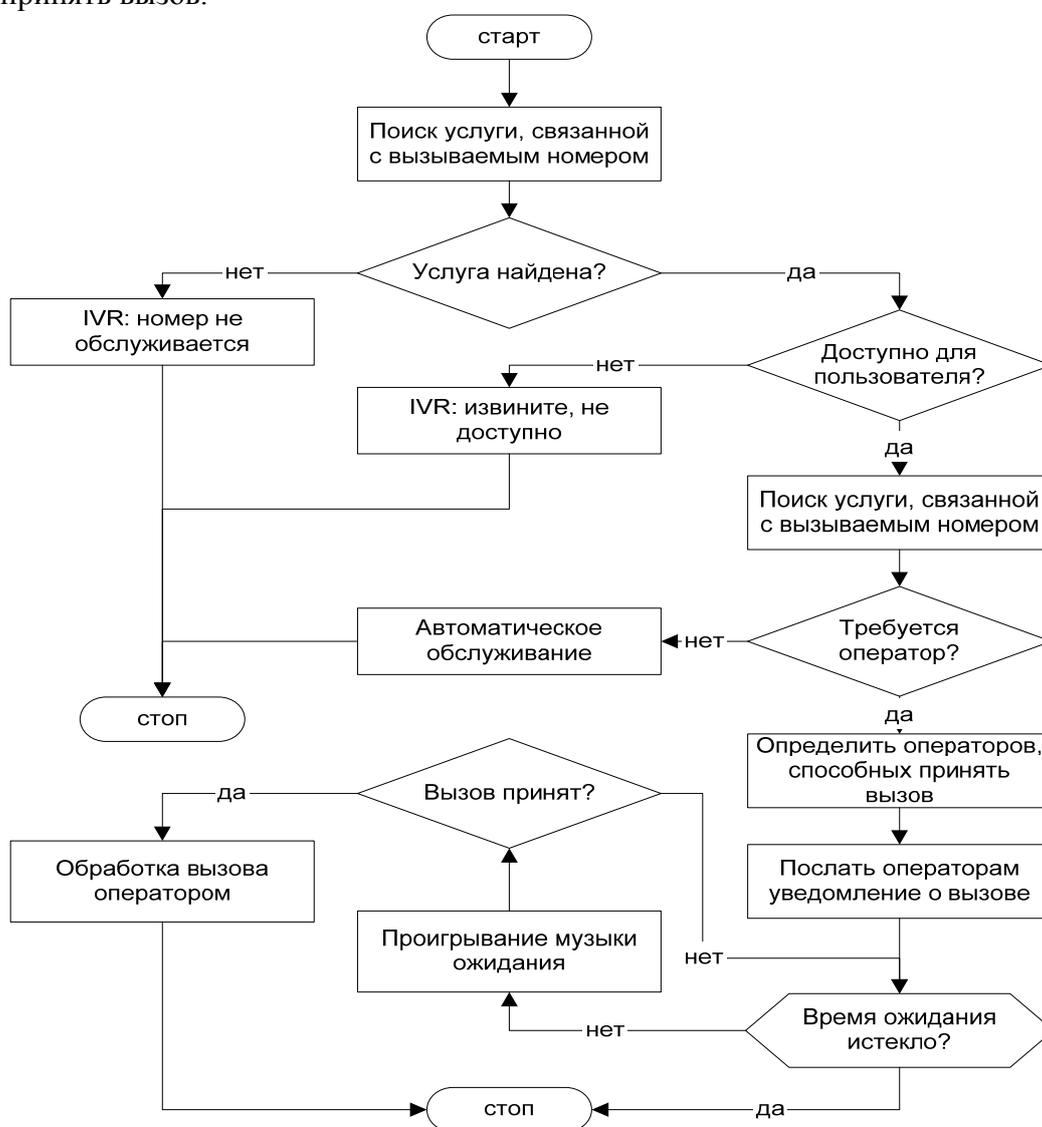


Рис. 4. Алгоритм обслуживания вызова

После принятия вызова оператором в его рабочее окно загружается форма, соответствующая данной услуге. Получив необходимую информацию от пользователя, оператор запускает процесс выполнения заказа.

Упрощенная диаграмма взаимодействия компонентов системы приведена на рис. 5. На диаграмме под пользователем понимается телефонный аппарат абонента; IP УАТС – коммутатор, на котором зарегистрирован обслуживаемый номер; САДУ – разрабатываемый сервер автоматизированного доступа к услугам, оператор – операторский телефон и web-приложение.

С внешним миром система взаимодействует посредством протоколов SIP (передача сигнальной информации), RTP (передача голосовых данных) и HTTP (загрузка web-страниц).

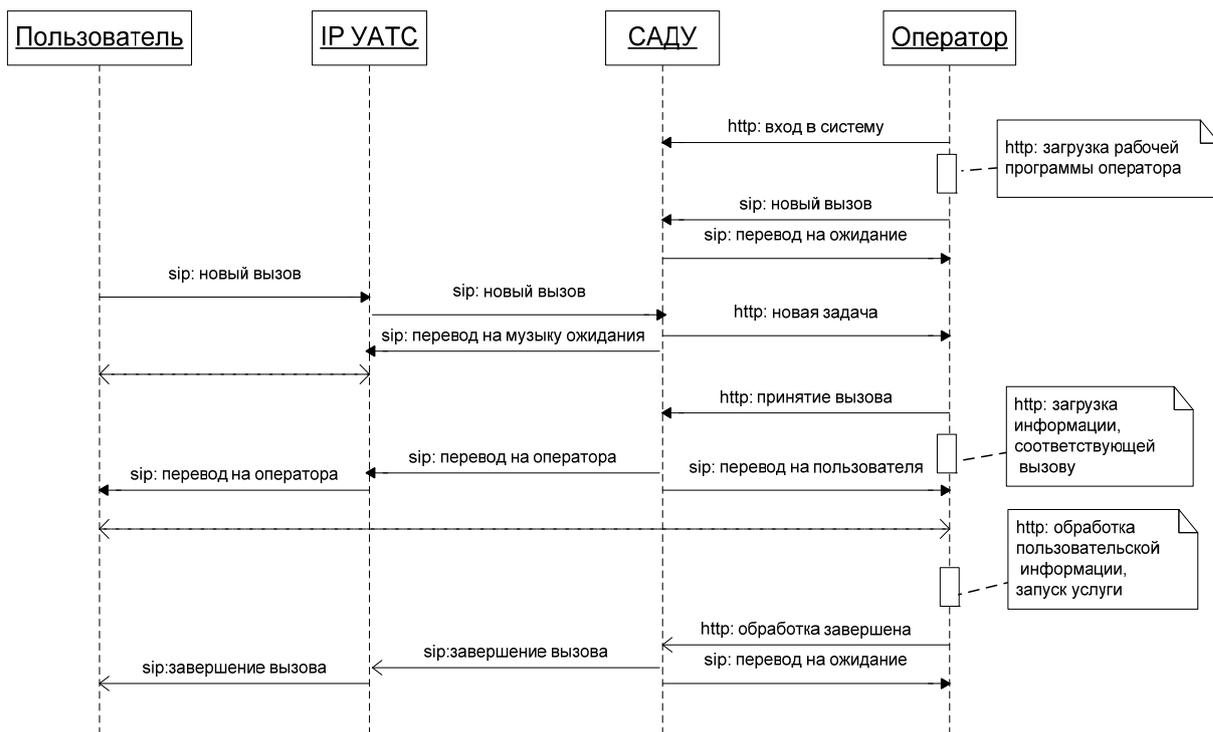


Рис. 5. Упрощенная диаграмма взаимодействия компонентов системы

Пользовательский интерфейс

К разрабатываемому интерфейсу были предъявлены следующие требования:

- доступность и простота управления, рассчитанная на непрофессионального пользователя;
- наглядность производимых действий;
- удобство в ориентации.

Кроме того, необходимо обеспечить возможность применения различных пользовательских операционных систем. Проанализировав эти требования, было принято решение в качестве клиентской части системы использовать Web-приложение, исполняющееся в Web-браузере. Такое приложение загружается каждый раз при использовании с Web-сервера. Такой подход гарантирует актуальность клиентской программы, не требует ее инсталляции и конфигурирования для каждой пользовательской ЭВМ. Физически пользовательское приложение представляет собой динамически изменяющуюся страничку HTML (DHTML) со вставками кода JavaScript и блоков стилей CSS.

На рис. 6 представлен внешний вид рабочего окна администратора. Данная консоль позволит администратору управлять всеми необходимыми сущностями системы. В первую очередь к ним относятся организации, услуги и сотрудники.

В верхней части консоли расположен заголовок, содержащий иконку проекта и информацию о текущем пользователе. При нажатии на соответствующие ссылки имеется возможность смены роли и выхода из системы.

В левом фрейме расположено меню. При выборе пункта меню раскрывается подменю и загружается соответствующая информация в правый фрейм, как показано для пункта Employee.

Правый фрейм состоит из двух блоков. Верхний содержит список сущностей, тип которых был выбран в меню. В данном случае отображен список сотрудников для данной организации. В списке два сотрудника. Информация о выбранной сущности отображается в нижнем блоке. Так мы можем просмотреть всю необходимую информацию о сотруднике turbine. При желании информация может быть изменена и сохранена.

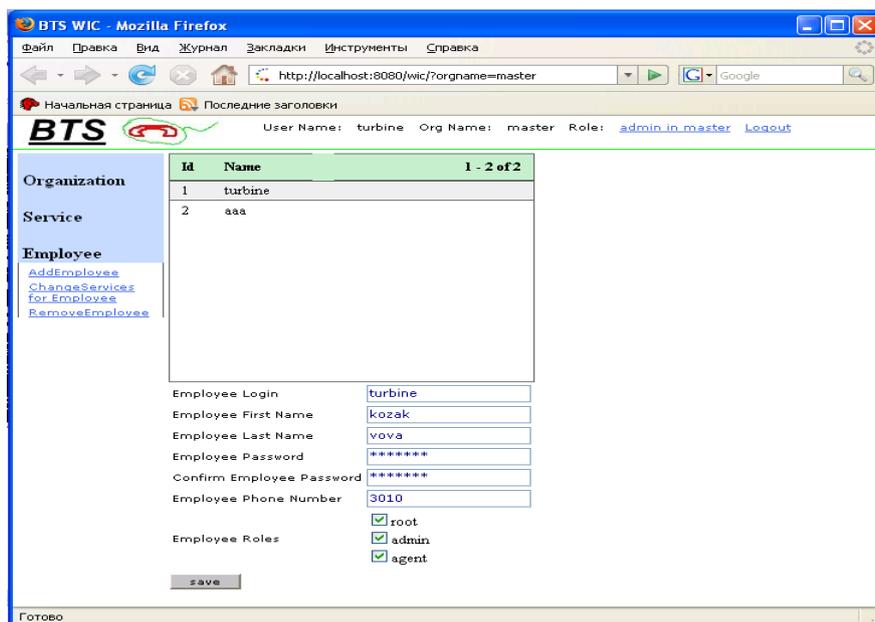


Рис. 6. Внешний вид рабочего окна администратора

Внешний вид рабочего окна оператора представлен на рис. 7. Это окно, помимо заголовка, содержит (снизу вверх):

- таблицу со списком вызовов и задач на обслуживание. Количество строк в таблице можно изменять. Выделенная строка и строка, над которой находится курсор, помечаются цветом;
- набор элементов управления. Включает меню, пункты которого динамически подсвечиваются при возможности совершения соответствующего действия. Помимо меню, имеется фильтр для типов услуг, поле ввода количества строк в таблице вызовов, индикатор количества задач в таблице и кнопка свертывания таблицы;
- фрейм для загрузки HTML-страницы, соответствующей вызову. Страница загружается после принятия вызова.

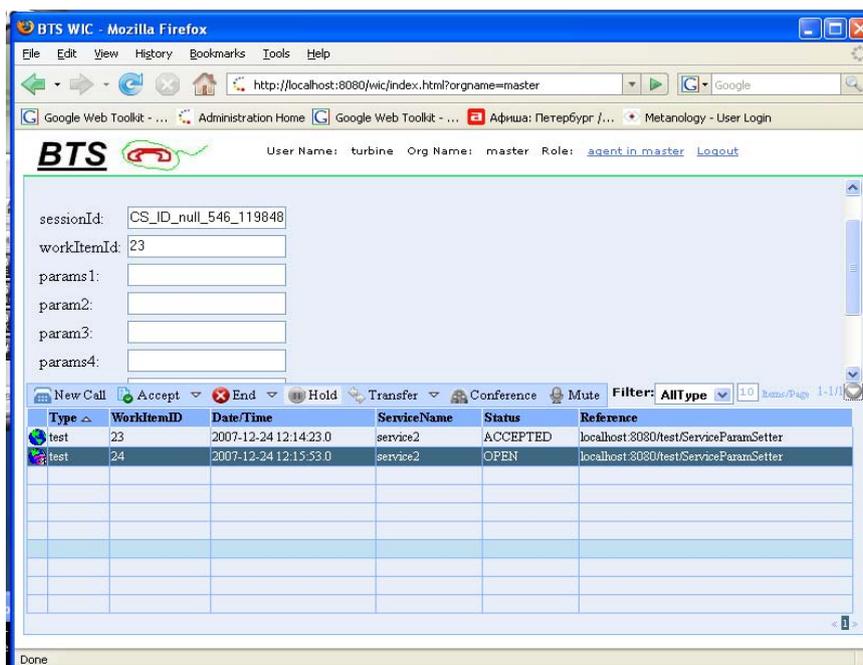


Рис. 7. Внешний вид рабочего окна оператора

Заключение

В рамках данного проекта развернута сеть IP-телефонии, настроены соответствующие службы и сервера, а также спроектирован и создан сервер автоматизированного доступа к услугам и клиенты к нему. Информация, требующая постоянного хранения, запоминается в реляционной базе данных. Не зависящие от пользовательской платформы клиентские приложения представляют из себя Web-приложения, исполняющиеся в Web-браузере. Все это является залогом быстрого и удобного развертывания системы для практического применения. Удобный и интуитивно понятный графический пользовательский интерфейс позволяет сократить время обучения персонала и повысить его производительность.

Литература

1. Гольфштейн А.Б., Гольфштейн Б.С. SOFTSWITCH. – СПб: БХВ - Санкт-Петербург, 2006. – 368 с.
2. Гольфштейн А.Б., Фрейкман В.А. Call-центры и компьютерная телефония. – СПб: БХВ - Санкт-Петербург, 2002. – 372 с.
3. Мадсен Л., Смит Д., Меггелен Д. Asterisk The Future of Telephony – O'Reilly, 2005. 404 с.
4. Монсон-Хейфел Р. Enterprise JavaBeans, 3-е издание / Пер. с англ. – СПб: Символ-Плюс, 2002. – 672 с.
5. Крейн Д., Паскарелло Э., Джеймс Д. Ajax в действии / Пер. с англ. – М.: ООО «И. Д. Вильямс», 2006. – 640 с.

СИСТЕМА КОНТРОЛЯ ЗА СИСТЕМОЙ ЧЕРЕЗ ЛОГ-ФАЙЛЫ

В.П. Соловьёв

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В настоящий момент осуществляется недостаточный контроль над системами после передачи заказчику. В данной статье рассматривается проблема контроля за системой через лог-файлы. Для дополнительного контроля вводится элемент искусственного интеллекта.

Введение

Создаваемое программное обеспечение (ПО) с каждым годом совершенствуется и становится более сложным. Сроки на создание ПО сокращаются. Все это сказывается на качестве программного кода. Требуется новые средства разработки программного обеспечения, новые возможности. Код программных продуктов со временем становится все сложнее. Как следствие, программисты чаще совершают ошибки. Для контроля за ошибками существуют разные средства прогнозирования и поиска ошибок. Однако даже самые передовые средства прогнозирования и поиска ошибок не способны полностью проконтролировать написание программного продукта.

Многие разработчики используют метод *логгирования* (от англ. – *logging*). Методология заключается в том, что программа пишет происходящие в ней процессы в лог-файл(ы): записывается информация о работе программы, в том числе и ошибки. В лог-файл записываются ключевые моменты работы программы, т.е. моменты выполнения ее логики работы. При возникновении ошибки или аварийном завершении программы происходит запись в лог-файл о причине. Поиск причины ошибки происходит в лог-файле. Поиск происходит различными способами, например, вручную. Способ поиска ошибок вручную имеет следующее преимущество: наглядно видно всю последовательность выполнения кода программы.

Недостатком способа является избыточное количество информации, содержащееся в лог-файле, что не позволяет быстро найти ошибку. Для просмотра лог-файла необходимо иметь лог-файл либо доступ к компьютеру, на котором работает ПО. Необходимо в кратчайшие сроки вернуть ПО в рабочее состояние и установить причину ошибки. Удаленное расположение лог-файла влияет на оперативность. Доступ к лог-файлу является необходимым, но не достаточным условием нахождения причины ошибки. Для устранения настоящих недостатков необходимо разработать специальную систему, которая автоматизирует ряд функций. Следует на протяжении всего времени работы ПО отслеживать его состояние через лог-файлы и создавать уведомления о происходящих ошибках.

Возможна ситуация, при которой появление ошибки не влечет запись в лог-файл. Такую ситуацию должен отслеживать особый элемент системы – элемент искусственного интеллекта (ИИ). Элемент ИИ следит за ходом выполнения программы, и сигнализирует о нестандартном поведении программы.

1. Описание проблемы

Человек в своей деятельности допускает ошибки. Количество допускаемых человеком ошибок зависит от многих факторов: компетенции, физического состояния человека, рабочего инструмента и др.

Деятельность программиста имеет особую специфику. Эту специфику необходимо учитывать при создании рабочего процесса. Как бы хорошо ни был налажен рабочий процесс, всегда существует вероятность появления ошибки во время рабочего процесса. Не всегда ошибку можно найти до того, как программный продукт реализован (передан заказчику). В этом случае ошибки проявляются уже во время пользования

программы. Для выявления и устранения ошибок существует много разных систем поиска и предотвращения ошибок. Их разделяют на два вида: во время создания ПО и во время использования ПО.

Встроенные статические системы слежения за ошибками в среде разработки программного обеспечения пригодны для поиска опечаток и неправильного использования правил ООП. Они не пригодны для отслеживания всех ошибок и часто дают ложные результаты. Динамические системы используются программистами, когда известны условия появления ошибки. Динамический поиск ошибки представляет собой пошаговое выполнение программы. При этом используются специальные метки (*breakpoints*).

Ко второму виду относятся системы ведения лог-файлов и встроенные системы. Они являются дополнительными средствами слежения за ошибками. Как правило, все средние и крупные программные проекты разрабатываются с теми или иными дополнительными средствами слежения за ошибками. Распространенным средством слежения за логическими ошибками является слежение за выполнением хода программы. Для этого используется метод, называемый *логгированием* (от англ. *Logging – регистрация, запись*). В общем системы логгирования работают одинаково: в коде разрабатываемого ПО указываются правила вывода в лог-файл. В лог-файл выводятся ключевые моменты в работе программного продукта. Следовательно, по лог-файлу можно проследить ход выполнения программы и обнаружить ошибку.

Формат записи в лог-файл настраивается. Настраиваются такие параметры, как время (число, месяц, год или миллисекунды), поток выполнения (*thread*), уровень важности записи и др.

Необходимость автоматизировать процесс контроля ошибок в лог-файле приводит к созданию системы контроля за ПО через лог-файл. Система должна автоматически оповещать об ошибке. Если ПО работает удаленно, необходимо удаленно оповещать, например, через электронную почту. Результаты анализа лог-файлов необходимо сохранять. Для этого подходит любая реляционная база данных. Чтобы детерминировать реакцию системы на найденные ошибки, она должна осуществлять поиск согласно шаблонам. Шаблоны должны настраиваться для каждого лог-файла или всех сразу. По найденным ошибкам система должна генерировать отчет. Имея отчет, можно сделать вывод по работе ПО.

Система должна понимать формат записи в лог-файлах. На протяжении всего времени работы ПО система следит за новыми записями в лог-файл. При наличии нескольких лог-файлов система должна следить за всеми. Система должна иметь возможность оповещения об ошибках удаленно. Система должна понимать формат записи в лог-файл. Формат записи в лог-файл должен находиться в конфигурационном файле, и быть конфигурируемым. Система не должна оповещала обо всех происходящих ошибках.

Разные ошибки генерируются в разных участках программы, которые могли быть написаны разными разработчиками. Каждый разработчик должен получать уведомление об ошибке, произошедшей в его части кода. Это должно настраиваться в конфигурационном файле. Следует иметь возможность группового оповещения об ошибке, если ответственность несет группа разработчиков.

Система должна иметь возможность самостоятельно принимать решения. Для этого в системе должен быть элемент искусственного интеллекта (ИИ). Элемент ИИ следить за ходом выполнения программы и строит дерево выполнения ПО. По дереву выполнения ПО делается вывод о состоянии ПО. Стандартное поведение ПО означает, что выполняется только «полезный» код и в нужной последовательности, т.е. код основной логики в определенной последовательности.

2. Концепция проекта и верхний уровень структуры

Система имеет распределенную структуру, благодаря которой возможно размещение каждой ее части физически в разных местах. Все необходимые настройки производятся в конфигурационных файлах. Система представляет собой: сервер, базу данных и агентов. Схема показана на рис. 2.1.

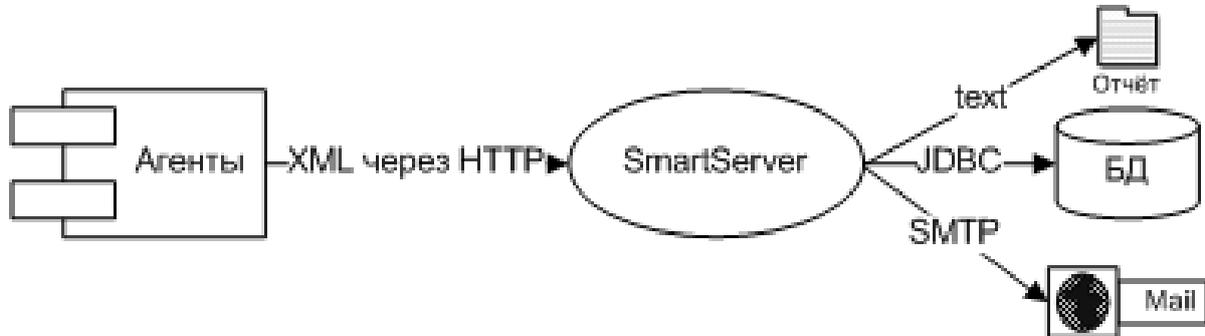


Рис. 2.1. Общая структура системы

Агенты – агенты, взаимодействующие с сервером. К системе можно подключить одного и более агентов. SmartServer – сервер, связующее звено между агентами, базой данных (БД) и почтой. Сервер получает от агентов сообщения в XML-формате, записывает сообщения в БД, создает отчеты и отправляет уведомления на почту (*mail*). Отправка уведомлений на почту происходит через SMTP-протокол. Настройка сервера происходит через конфигурационный файл. Отчеты сервер записывает в текстовом виде или xml-формате. БД – база данных, которая служит для хранения событий. Mail – электронная почта, которая служит для рассылки писем об ошибках.

Сервер записывает данные, полученные от агентов, в базу данных через JDBC (Java DataBase Connectivity) и отправляет уведомления по почте через протокол SMTP (Simple Mail Transfer Protocol)

Каждый агент является независимым. Между Лог-агентом и Смарт-агентом сделана зависимость, так как они обрабатывают общие лог-файлы. Взаимодействие между Лог-агентом и Смарт-агентом показано на рис. 2.2.

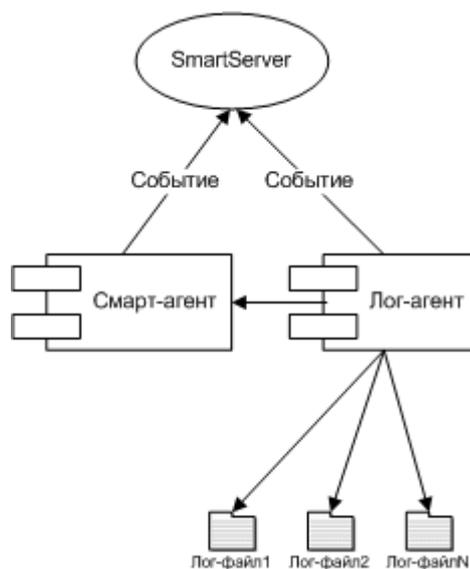


Рис. 2.2. Взаимодействие агентов

Несколько лог-файлов (Лог-файл1, Лог-файл2, Лог-файлN) читаются Лог-агентом. Лог-агент передает обработанную информацию Смарт-агенту. Каждый из агентов выполняют свое предназначение: Лог-агент ищет ошибки в соответствии с шаблонами, Смарт-агент играет роль элемента ИИ.

3. Конфигурирование системы

Конфигурационные файлы имеют XML-формат. XML – универсальный формат [1]. Имеется два конфигурационных файла: Server.xml и Agents.xml. В Server.xml описана конфигурация для сервера. В Agents.xml описана конфигурация для агентов. Конфигурирование системы представлено на рис. 3.1.

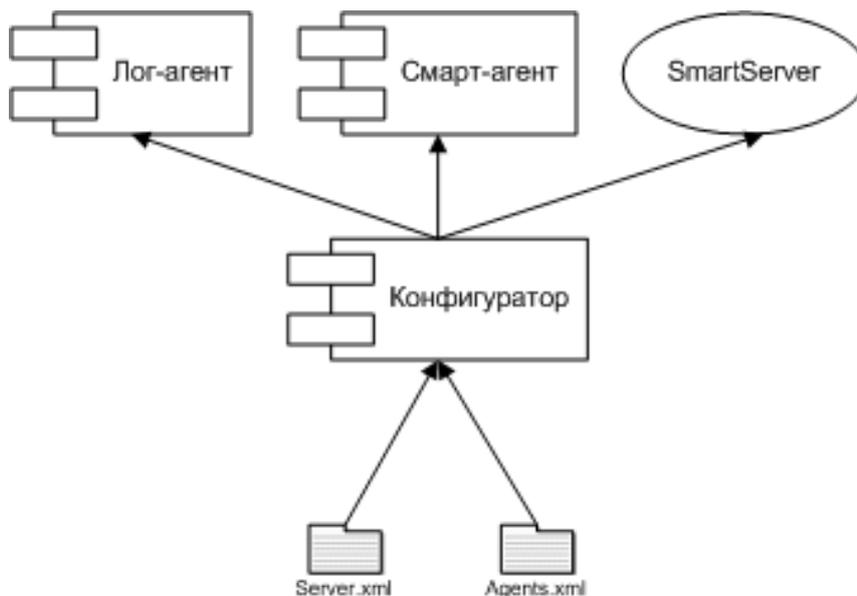


Рис. 3.1. Конфигурация системы

Конфигуратор создает агентов и сервер по конфигурационным файлам.

4. Элемент ИИ

Программное обеспечение записывает последовательность своей работы в лог-файлы. Лог-файлов может быть несколько. Их разделяют по предназначению. Предназначение зависит от модуля программы или от типа события. Лог-агент взаимодействует со Смарт-агентом с целью получения считанной строки. Лог-агент считывает и анализирует строку. Лог-агент передает считанные данные Смарт-агенту, который выполняет свою функцию. Каждый лог-файл контролируется отдельно. Взаимодействие агентов представлено на рис. 4.1.

Лог-агент – агент системы, элемент прямого анализа Лог-файлов. Смарт-агент – агент системы, элемент искусственного интеллекта (ИИ). Лог-файл1, Лог-файл2, Лог-файл3 – лог-файлы. Каждый агент работает в отдельном потоке. Это необходимо для того, чтобы каждая часть программы работала независимо от других частей. Запущенные в разных потоках элементы системы не влияют на работу других элементов.

Чтобы избежать закливания при обращении к лог-файлу, поток, в котором работает Лог-агент, «засыпает» на некоторый интервал времени.

Считывание из лог-файла осуществляется построчно Лог-агентом. Каждая считанная строка, проверяется на схожесть с шаблонами. Если обнаружено сходство счи-

танной строки с шаблоном, то генерируется событие ошибки. Данные считанной строки передаются Смарт-агенту, который обрабатывает их относительно своей функциональности.

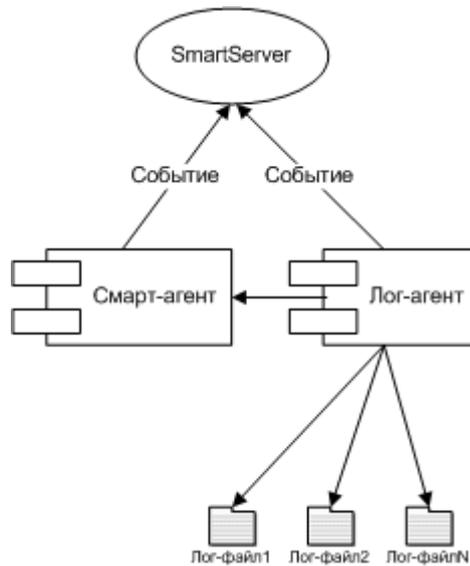


Рис. 4.1. Мониторинг лог-файлов

Каждый лог-файл контролируется в отдельном потоке. Это позволяет повысить скорость работы программы и обезопасить контроль файлов. Возникновение ошибки чтения в одном из файлов не влияет на процесс чтения других.

Пример записи в лог-файле:

32 дек 2007 13:57:12 main ERROR [smart.server.SmartLogServer.start()] Server failed to start
32 дек 2007 13:57:12 – время записи.

main – поток записи в лог-файл.

ERROR – уровень значимости. Другие возможные значения: *DEBUG*, *WARN*, *FATAL*.

[smart.server.SmartLogServer.start()] – указание места, где произошла ошибка.

Server failed to start – описание ошибки.

Имя, причина и стек ошибки в программе:

```
org.mortbay.util.MultiException[com.smart.SmartLogException: Failed to open dataStore,
java.net.BindException: Cannot assign requested address: JVM_Bind]
  at org.mortbay.http.HttpServer.doStart(HttpServer.java:686)
  at org.mortbay.util.Container.start(Container.java:72)
  at com.smart.server.SmartLogServer.start(SmartLogServer.java:60)
  at com.smart.server.ServerLauncher.main(ServerLauncher.java:62)
  at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
  at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:39)
  at
  sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:25)
  at java.lang.reflect.Method.invoke(Method.java:597)
  at com.intellij.rt.execution.application.AppMain.main(AppMain.java:90)
```

Так выглядит наиболее детальная информация, которая анализируется разработчиками, отладчиками или др.

Система по заданным шаблонам отслеживает ошибки и уведомляет о них на почту. Сообщение о случившемся сохраняется в БД.

Существует ряд ошибок, информация о которых не регистрируется в лог-файле. Следующий шаг автоматизации – самостоятельность системы в контроле за происходящими ошибками [2, 3]. Система автоматически обнаруживает новые записи и сигналами

лизирует об ошибках. По обнаруженным записям и ошибкам собирается статистика. По собранной статистике строится дерево. Ветви дерева – этапы работы программы [4]. Анализ статистики показывает, насколько часто после одного события происходит другое. По частоте встречаемости одного события после другого система вычисляет вероятность. В зависимости от того, насколько мало значение вероятности наступления события, система генерирует событие ошибки или предупреждения. Пороги срабатывания системы наглядно представлены в табл. 4.1.

вероятность события (p)	$P \leq 0.01$	$0.01 < p \leq 0.1$	$0.1 < p$
действие системы	генерация ошибки	генерация предупреждения	ничего

Таблица 4.1. Зависимость генерации события от вероятности встречаемости

Иллюстрация алгоритма Смарт-агента приведена на рис. 4.2

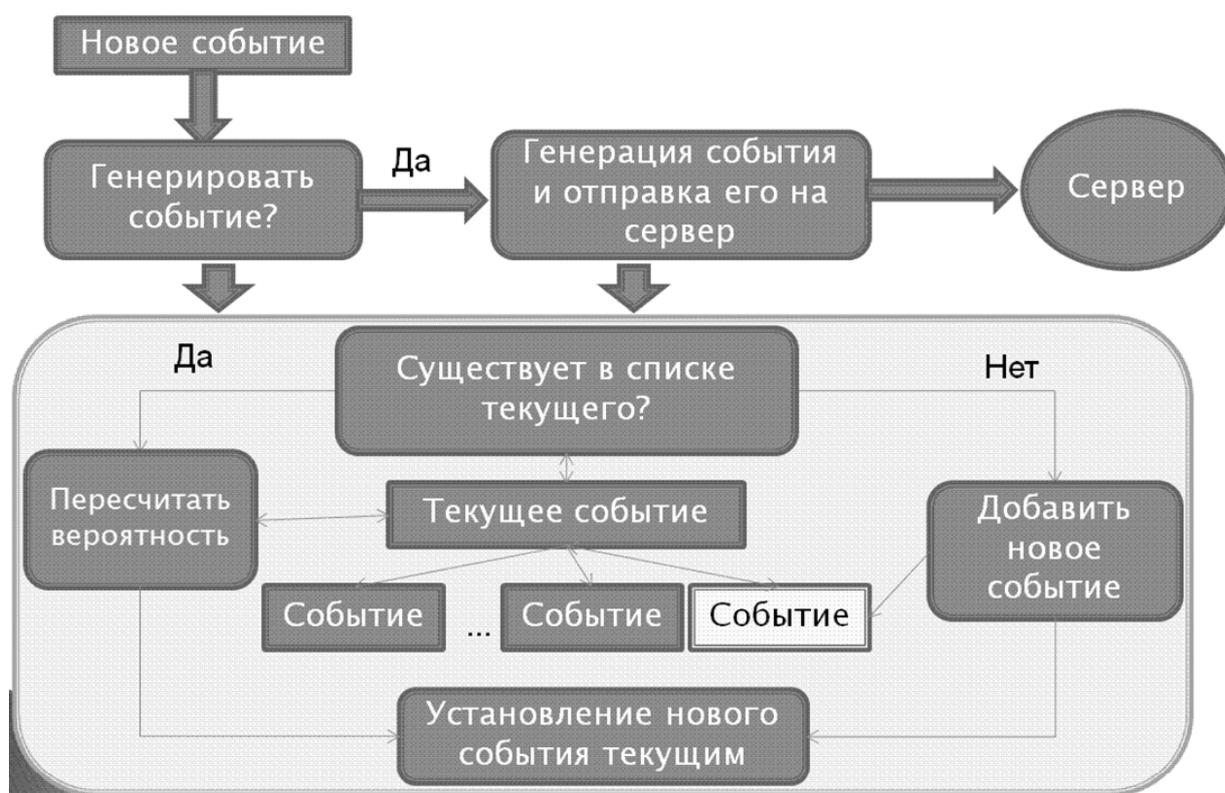


Рис. 4.2. Иллюстрация главного алгоритма

На рис. 4.2 представлен главный алгоритм Смарт-агента. Смарт-агент получает данные от Лог-агента и обрабатывает их. Вычисляется вероятность наступления нового события после текущего. Если нового события нет среди списка событий текущего события, событие добавляется в список. Полученная вероятность сравнивается с пороговыми значениями (см. табл. 4.1). Выполняется соответствующее действие системы. Проверяется наименее вероятный случай, если он наступает, то генерируется событие, и другие пороги срабатывания не учитываются. Проверяется, есть ли новое событие в списке текущего события. Если есть, происходит пересчет вероятностей. Если нет, новое событие добавляется в список событий текущего события, и пересчитываются вероятности. Новое событие устанавливается текущим.

Пороги срабатывания определяются эмпирическим путем. Наиболее правильные пороговые значения основаны на прошлой статистике ПО. Статистика должна быть собрана во время безошибочной работы программы. Пороги предупреждения и ошибки должны выбираться по наименее вероятной записи статистики. Более длительный период эмпирических данных обеспечит более точные результаты работы системы.

Наглядное представление дерева событий представлено на рис. 4.3.

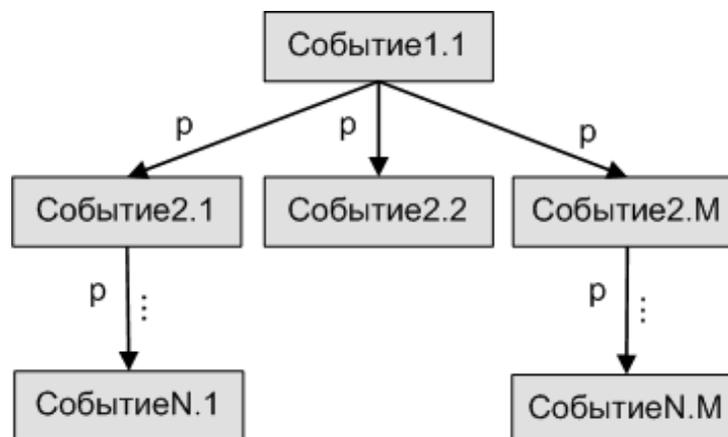


Рис. 4.3. Дерево событий

Здесь EventM.N – событие, которое создается по добавленным записям в лог-файл; p – вероятность наступления события. Сумма вероятностей наступления события в одном слое равна единице:

$$\sum_{i=1}^M p(\text{EventM.i}) = 1,$$

где M – номер слоя, N – ширина слоя.

5. Заключение

В статье рассмотрена проблема контроля за программным обеспечением, через лог-файлы. Описан механизм функционирования системы. Отличительной особенностью данной системы является наличие встроенного элемента искусственного интеллекта, что дает ей преимущество относительно аналогов. Тестирование системы продемонстрировало способность системы к обнаружению нестандартного поведения программного обеспечения через лог-файлы.

Систему можно использовать в качестве анализатора лог-файлов – как основного, так и дополнительного. Система не оказывает влияния на другие анализаторы лог-файлов, которые работают параллельно.

Литература

1. Холзнер С. XML Энциклопедия, 2-е издание. – Питер, 2004. – 1100 с.
2. Осовский С. Нейронные сети для обработки информации. – Изд. Финансы и статистика, 2004. – 344 с.
3. Рассел С., Норвинг П. Искусственный интеллект. Современный подход. 2-е издание. – 2007. – 1408 с.
4. М. Тим Джонс. Программирование искусственного интеллекта в приложениях. – 2006. – 312 с.

РАЗРАБОТКА И ИЗГОТОВЛЕНИЕ ЭЛЕКТРОННОГО УСТРОЙСТВА ДЛЯ ПРОВЕДЕНИЯ ОПРОСОВ

И.Н. Архипов (Лицей №344 Невского района Санкт-Петербурга)
Научный руководитель – В.В. Илларионова

Разработано и изготовлено электронное программируемое устройство, которое можно использовать для проведения экспресс-опросов. Устройство представляет собой печатную плату с установленными на ней AVR контроллером ATmega8, знакосинтезирующим индикатором и другими электронными компонентами, обеспечивающими работоспособность устройства в соответствии с поставленной задачей.

Разработка и создание устройства включают в себя следующие этапы.

1. Разработка программного кода, в соответствии с которым контроллер работает в устройстве. Создан в среде CodeVisionAVR, которая представляет собой компилятор высокого уровня, предназначенный для создания программного кода для контроллеров типа AVR.
 - 1.1. Создание алгоритма работы программы.
 - 1.2. Написание программы на языке «Си» и ее компиляция.
2. Проектирование принципиальной схемы и подбор компонентов. Принципиальная схема устройства представлена на рис. 1.

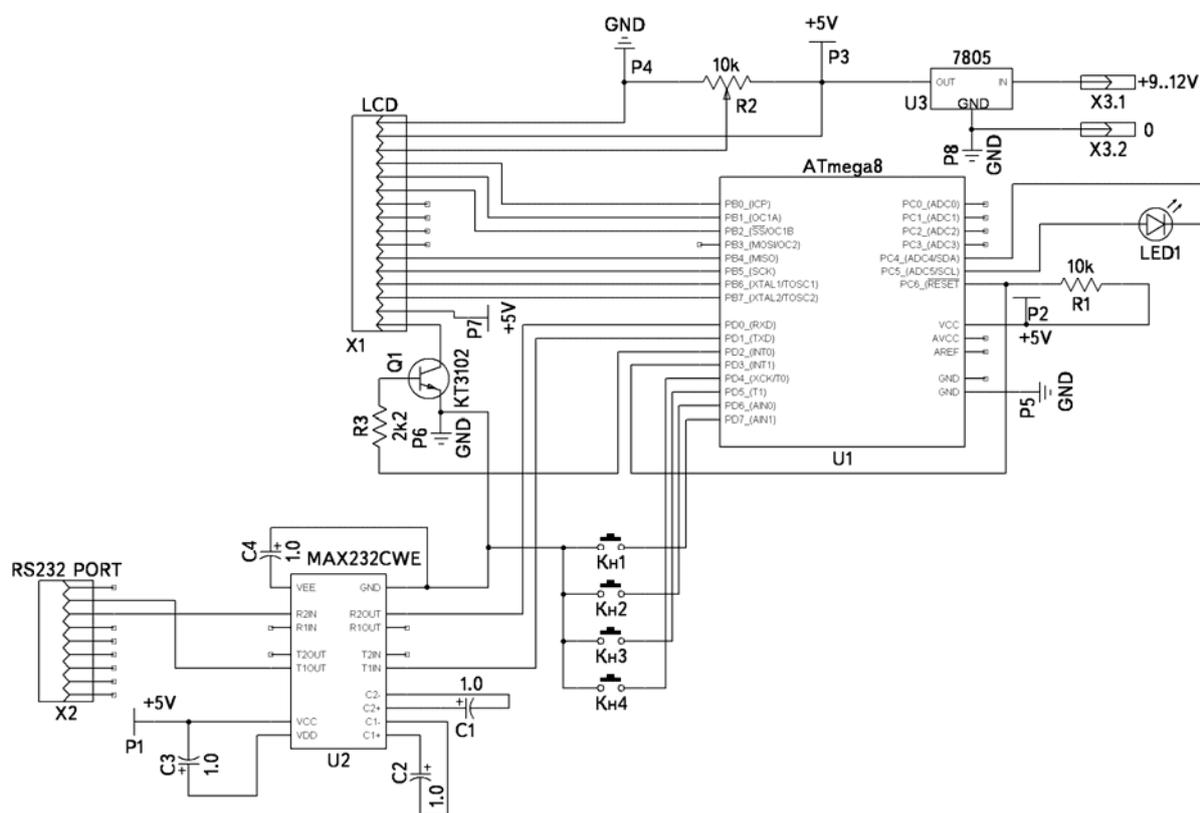


Рис. 1. Принципиальная схема

3. Контроллеры AVR включают в себя несколько семейств. В таблицах, представленных ниже, приводятся характеристики наиболее распространённых и часто используемых семейств Tiny и Mega.

Название	Flash ROM	EEPROM	I/O	Внутренний RC-генератор	Напряжение питания, V _{cc}	Тактовая частота, МГц	Тип корпуса
ATtiny13	1 KB	64 B	6	Есть	1.8-5.5, 2.7-5.5	0-10, 0-20	DIP8, SOIC8, SSOIC8, MLF20
ATtiny24	2 KB	128 B	12	Есть	1.8-5.5, 2.7-5.5	0-10, 0-20	DIP14, SOP14, MLF20
ATtiny44	4 KB	256 B	12	Есть	1.8-5.5, 2.7-5.5	0-10, 0-20	DIP14, SOP14, MLF20

Таблица 1. Характеристики контроллеров Tiny

Название	Flash ROM	EEPROM	I/O	Интерфейсы	Внутренний RC-генератор	Сторожевой таймер	Напряжение питания, V _{cc}	Тактовая частота, МГц	Тип корпуса
ATmega48	4 KB	256 B	23	UART, 2 SPI, I2C	Есть	Есть	1.8-5.5	0-10, 0-20	PDIP28 TQFP32 MLF32
ATmega8	8 KB	512 B	23	UART, SPI, I2C	Есть	Есть	2.7-5.5 4.0-5.5	0-8, 0-16	DIP28, TQFP32 MLF32
ATmega16	16 KB	512 B	32	UART, SPI, I2C	Есть	Есть	2.7-5.5 4.5-5.5	0-8, 0-16	DIP40, TQFP44

Таблица 2. Характеристики контроллеров Mega

Анализ основных параметров микроконтроллеров показал, что могут использоваться микроконтроллеры семейства Mega. Контроллеры семейства Tiny в данном случае применять нельзя, так как они имеют мало портов ввода-вывода и не смогут управлять знакосинтезирующим индикатором и считывать состояние кнопок (минимум 7 ка-

налов ввода-вывода для дисплея и 4 для кнопок), или мало Flash-памяти, и если понадобится дополнить программу, то ее может не хватить. Было решено использовать контроллер в стандартном PDIP корпусе, а это значит, что ATmega16 не поместится в корпус устройства, и надо выбирать между ATmega48 и ATmega8. ATmega8 имеет больше Flash-памяти, чем ATmega48, по этой причине я выбрал микроконтроллер ATmega8.

Для правильной связи с компьютером (RS-232) необходим драйвер порта – микросхема MAX232.

- 3.1. Моделирование электронного устройства в программе-симуляторе Proteus.
- 3.2. Создание проверенной схемы в пакете программ DipTrace.
- 3.3. Преобразование принципиальной схемы в плату и ее трассировка.
4. Оптимизация как схемотехнического решения, так и программного кода с целью экономии ресурсов контроллера, увеличения скорости исполнения программы, уменьшения габаритов устройства. Проверка правильности монтажной схемы устройства, отсутствия наложения компонентов, наличия необходимых зазоров между дорожками токопроводящих цепей.
5. Изготовление печатной платы методом фототравления.
 - 5.1. Вырезание из листа стеклотекстолита заготовки нужных размеров.
 - 5.2. Зачистка и обезжиривание фольгированного стеклотекстолита.
 - 5.3. Покрытие светочувствительным лаком (фоторезистом) фольгированной поверхности платы.
 - 5.4. Печать фотошаблона (дорожек) на прозрачной пленке.
 - 5.5. Экспонирование заготовки с наложенным фотошаблоном с помощью УФ излучения (например, лампа дневного света).
 - 5.6. Проявка лака (растворение засвеченных участков в щелочной среде, напр. NaOH).
 - 5.7. Травление печатной платы (химическое удаление медной фольги, не защищённой лаком).
 - 5.8. Сверловка отверстий, необходимых для установки всех компонентов и элементов устройства на плате.
6. Монтаж всех электронных компонентов на печатной плате пайкой и установка в корпус.



Рис. 2. Внешний вид устройства

В процессе работы над устройством были самостоятельно изучены следующие темы.

- Программирование контроллеров семейства AVR в среде CodeVisionAVR.
- Моделирование электронных схем в специализированной программе Proteus.

- Проектирование печатной платы с помощью пакета программ DipTrace.
- Изготовление печатных плат в домашних условиях с помощью фоторезиста.

Параллельно с этим устройством был собран программатор AVR контроллеров для возможности загрузки кода программы (прошивки) в память контроллера.

При создании устройства использована информация, размещенная на специализированных сайтах в сети Интернет.

В результате проделанной работы сконструировано работоспособное устройство. Оно запрограммировано для тестирования по информатике. Внешний вид устройства представлен на рис. 2.

Литература

1. Евстигнеев А.В. Микроконтроллеры AVR семейств Tiny и Mega фирмы «ATMEL». – М.: Издательский дом «Додэка-XXI», 2004.
2. Голубцов М.С. Микроконтроллеры AVR: от простого к сложному. – М.: СОЛОН-Пресс, 2003.
3. Very Reasonable Technological Pages [Электронный ресурс]. – Режим доступа: <http://www.vrtp.ru>
4. Магазин «Микроника» в Санкт-Петербурге [Электронный ресурс]. – Режим доступа: <http://www.micronika.ru>

**МЕТОДИКА ОЦЕНКИ ЗАЩИЩЕННОСТИ
КОНФИДЕНЦИАЛЬНОЙ ИНФОРМАЦИИ, ОБРАБАТЫВАЕМОЙ
ОСНОВНЫМИ ТЕХНИЧЕСКИМИ СРЕДСТВАМИ, ОТ ЕЕ УТЕЧКИ
ЗА СЧЕТ ПОБОЧНЫХ ЭЛЕКТРОМАГНИТНЫХ ИЗЛУЧЕНИЙ, ПРИ
ИСПОЛЬЗОВАНИИ СРЕДСТВ АКТИВНОЙ ЗАЩИТЫ**

**Е.А. Голунчикова (Московский государственный технический университет
им. Н.Э. Баумана), А.Л. Липатов
Научный руководитель – д.т.н., профессор Ю.А. Гатчин**

В настоящей работе приводится методика оценки защищенности информации, содержащей сведения конфиденциального характера, обрабатываемой техническими средствами, от ее утечки по каналам побочных электромагнитных излучений. Предложенная методика позволит проводить оценку защищенности информации от утечки по каналам побочных электромагнитных излучений при использовании средств активной защиты информации.

Введение

Защита конфиденциальной информации от утечки по каналам побочных электромагнитных излучений (ПЭМИ) является основным направлением в обеспечении информационной безопасности (ИБ) информационно-телекоммуникационных систем (ИТС) [1]. Оценка защищенности основных технических средств ИТС, обрабатывающих конфиденциальную информацию, производится инструментально-расчетным методом в процессе специальных исследований на побочные электромагнитные излучения и наводки (ПЭМИН) [2, 3]. Специальные исследования проводятся в соответствии с методикой, изложенной в нормативно-методическом документе «Временная методика оценки защищенности основных технических средств и систем...». В результате специальных исследований определяется, являются ли основные технические средства защищенными от утечки за счет ПЭМИ.

В настоящей работе не содержится никакой конфиденциальной информации и на основе открытого математического аппарата, известного из теории распространения радиоволн, предлагается методика, которая позволит проводить оценку защищенности информации от утечки по каналам ПЭМИ при использовании средств активной защиты информации.

Методика оценки защищенности

«Временная методика оценки защищенности основных технических средств и систем...» предусматривает расчет зоны R_2 , т.е. требуемого радиуса контролируемой зоны (КЗ) вокруг основных технических средств ИТС, в которых осуществляется обработка конфиденциальной информации. При обеспечении вокруг основных технических средств КЗ, равной или большей R_2 , считается, что ИТС является защищенной от утечки конфиденциальной информации за счет ПЭМИ [4].

На практике случается, что радиус КЗ значительно меньше зоны R_2 , и организационными мерами обеспечить защищенности информации от утечки за счет ПЭМИ не удастся. В таких случаях требуется обеспечить либо экранирование помещения, в котором располагаются незащищенные основные технические средства ИТС, либо обеспечить установку средств активной защиты (генераторов пространственного шума типа

«ЛГШ-501», «Гном-3» и т.д.). Однако «Временная методика оценки защищенности основных технических средств и систем...» не предполагает оценки эффективности принятых мер по защите информации. Теоретически возможно рассчитать отношение сигнал/шум на границе КЗ. Но само по себе отношение нам ничего не дает: его нужно сравнивать с нормами, которые в неявном виде, но приведены во «Временной методике оценки защищенности основных технических средств и систем...» [4].

Согласно «Временной методике оценки защищенности основных технических средств и систем...» возможные расстояния распространения информативного сигнала от основных технических средств ИТС для каждой его спектральной составляющей рассчитывается по формуле

$$R = \frac{L}{\Pi \sqrt{\frac{K E_{ш}}{E}}}, (м)$$

В данной формуле параметры Π и L зависят от условий проведения измерений и свойств исследуемых основных технических средств (частот излучения). Параметр K – величина, зависящая от того, содержат ли исследуемые основные технические средства в своем составе монитор. А так как этот коэффициент непосредственно присутствует в отношении сигнал/шум, то можно сделать вывод, что это не что иное, как норма отношения сигнал/шум, при выполнении которой основные технические средства ИТС считаются защищенными. Значение K является закрытым, приведено в «Временной методике оценки защищенности основных технических средств и систем...» и выбирается в зависимости от того, имеются ли в составе основных технических средств ИТС [4].

Следовательно, можно предположить, что при выполнении отношения сигнал/шум на границе КЗ с значениями K меньшими, чем значения, приведенные во «Временной методике оценки защищенности основных технических средств и систем...», основные технические средства ИТС будут защищены от утечки за счет ПЭМИ.

Теперь требуется рассчитать фактическое отношение сигнал/шум на границе КЗ. При установке средств активной защиты в непосредственной близости от защищаемого основного технического средства можно считать, что коэффициент затухания информативного сигнала и шума при распространении до границы КЗ будет одинаковым. А, следовательно, отношение сигнал/шум будет постоянным на всей трассе распространения сигнала.

Следовательно, в таком случае достаточно измерить и рассчитать уровень информативного сигнала, а также измерить и рассчитать уровень шумовой помехи на каждой из частот ПЭМИ. Далее, по формуле требуется определить отношение сигнал/шум на каждой из частот:

$$\Delta_i = \frac{E_{c_i}}{E_{ш_i}},$$

где E_c и $E_{ш}$ – напряженности поля информативного сигнала и помех на i -й частоте в мкВ/м.

Теперь полученные величины требуется сравнить с нормами. Если на всех частотах обнаруженных ПЭМИ выполняется норма отношения сигнал/шум, то основные технические средства ИТС можно считать защищенным. Если хотя бы на одной из частот норма не выполняется – основные технические средства ИТС не защищены.

Заключение

Предложенная методика оценки защищенности конфиденциальной информации, обрабатываемой техническими средствами ИТС, от ее утечки по каналам ПЭМИ позволит учитывать на объекте защиты принятые меры по защите информации. Данный

факт приведет к созданию более эффективных систем защиты информации, обеспечивающих ИБ ИТС от актуальных угроз, реализуемых с использованием уязвимостей ИТС технического характера.

Литература

1. Концепция защиты средств вычислительной техники и автоматизированных систем от несанкционированного доступа к информации: Руководящий документ // Сборник руководящих документов по защите информации от несанкционированного доступа. – М.: Гостехкомиссия России, 1998.
2. Галкин А.П. Защита технических каналов связи предприятий и от несанкционированного доступа к информации: учеб. пособие / А.П. Галкин, В.С. Эмдин; М-во Рос. Федерации по связи и информатизации, Санкт-Петерб. гос. ун-т телекоммуникаций им. проф. М.А. Бонч-Бруевича. – СПб: СПбГУТ, 2003. – 99 с.
3. Хорев А.А. Защита информации от утечки по техническим каналам. Часть 1. Технические каналы утечки информации: Учебное пособие. М.: Гостехкомиссия России, 1998.
4. Временная методика оценки защищенности основных технических средств и систем...: Нормативно-методический документ // Сборник временных методик оценки защищенности конфиденциальной информации от утечки по техническим каналам. – М.: Гостехкомиссия России, 2002.

ПРОБЛЕМЫ БЕЗОПАСНОСТИ В БЕСПРОВОДНЫХ ЛВС IEEE 802.11

Д.С. Туранцев

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье рассматриваются вопросы безопасности в беспроводных локальных сетях стандарта IEEE 802.11, а также уязвимости протоколов, которые используются в данных сетях.

Введение

С момента ратификации стандарта IEEE 802.11b в 1999 году беспроводные ЛВС получили широкое распространение. Сегодня их можно встретить во многих офисах, конференц-залах, на промышленных складах, в школьных классах, кафе в деловой части мегаполиса. Беспроводные ЛВС стандарта IEEE 802.11b представляют собой ряд новых проблем для администраторов сетей и систем безопасности. В отличие от проводных сетей Ethernet, беспроводные ЛВС стандарта IEEE 802.11b используют общедоступный радиоканал для связи с абонентами. Этот факт лежит в основе целого ряда новых сложных проблем, решение которых потребовало дополнений к стандарту IEEE 802.11.

Средства обеспечения безопасности, предусмотренные спецификацией IEEE 802.11 и применимые также к 802.11b, 802.11a, 802.11g, подверглись тщательно-му анализу и серьезной критике. Аналитиками были выявлены и продемонстрированы серьезные уязвимости в определенных стандартом механизмах аутентификации (authentication), обеспечения конфиденциальности (privacy) и целостности (integrity) данных.

На данный момент большинство фирм и предприятий все больше внимания уделяют использованию непосредственно Wi-Fi-сетей. Обусловлено это удобством, мобильностью и относительной дешевизной при связи отдельных офисов и возможностью их перемещения в пределах действия оборудования. В Wi-Fi-сетях применяются сложные алгоритмические математические модели аутентификации, шифрования данных, контроля целостности их передачи, что позволяет быть относительно спокойным за сохранность данных при использовании данной технологии. Однако такая безопасность относительна, если не уделять должного внимания настройке беспроводной сети. К данному моменту уже существует список «стандартных» возможностей, которые может получить хакер при халатности в настройке беспроводной сети:

- доступ к ресурсам локальной сети;
- прослушивание, воровство трафика (имеется в виду непосредственно Интернет-трафик);
- искажение проходящей в сети информации;
- внедрение поддельной точки доступа;
- рассылка спама от имени вашей сети.

1. Аутентификация в беспроводных локальных сетях стандарта IEEE 802.11

Беспроводные ЛВС, ввиду их ширококвещательной природы, требуют реализации дополнительных механизмов для:

- *аутентификации абонентов* (user authentication) с целью предотвращения несанкционированного доступа к сетевым ресурсам;
- обеспечения *конфиденциальности данных* (data privacy) с целью обеспечения целостности и защиты при передаче по общедоступному радиоканалу.

Стандарт IEEE 802.11 предусматривает два механизма аутентификации беспроводных абонентов: открытую аутентификацию (open authentication) и аутентификацию с общим ключом (shared key authentication). Также широко используются два других механизма, а именно назначение идентификатора беспроводной ЛВС (Service Set Identifier, SSID) и аутентификация абонента по его MAC-адресу (MAC address authentication). Ниже рассмотрены перечисленные механизмы и присущие им недостатки.

Ключи шифрования WEP (Wired Equivalent Privacy) могут быть использованы в качестве своего рода механизма ограничения доступа, поскольку абонент, не обладающий корректным WEP-ключом, не сможет ни принять, ни отправить данные в беспроводную ЛВС. Технология шифрования WEP стандарта IEEE 802.11 оперирует ключами длиной 40 либо 104 бита. Технология WEP и присущие ей недостатки рассмотрены в последующих разделах.

Идентификатор беспроводной ЛВС (Service Set Identifier, SSID). SSID представляет собой атрибут беспроводной ЛВС, позволяющий логически отличать сети друг от друга. В общем случае, абонент беспроводной сети должен задать у себя соответствующий SSID для того, чтобы получить доступ к требуемой беспроводной ЛВС. SSID ни в коей мере не обеспечивает конфиденциальность данных, равно как и не аутентифицирует абонента по отношению к точке радиодоступа беспроводной ЛВС.

Аутентификация абонента в IEEE 802.11. Аутентификация в стандарте IEEE 802.11 ориентирована на аутентификацию абонентского устройства радиодоступа, а не конкретного абонента как пользователя сетевых ресурсов. Стандарт предусматривает два режима аутентификации: открытую и с общим ключом.[5]

Процесс аутентификации абонента беспроводной ЛВС IEEE 802.11 состоит из следующих этапов (рис. 1).

1. Абонент (Client) посылает фрейм *probe request* во все радиоканалы.
2. Каждая точка радиодоступа (access point, AP), в зоне радиовидимости которой находится абонент, посылает в ответ фрейм *probe response*.
3. Абонент выбирает предпочтительную для него точку радиодоступа и посылает в обслуживаемый ею радиоканал запрос на аутентификацию (*authentication request*).
4. Точка радиодоступа посылает подтверждение аутентификации (*authentication reply*).
5. В случае успешной аутентификации абонент посылает точке радиодоступа *association request*.
6. Точка радиодоступа посылает в ответ фрейм *association response*.
7. Абонент может теперь осуществлять обмен пользовательским трафиком с точкой радиодоступа и проводной сетью.

Ниже детально описаны процессы, происходящие на каждом из этапов.



Рис. 1. Процесс аутентификации абонента IEEE 802.11

Обмен фреймами Probe Requests, Probe Responses

При активизации беспроводной абонент начинает поиск точек радиодоступа в своей зоне радиовидимости с помощью управляющих фреймов probe request. Фреймы

probe request посылаются в каждый из радиоканалов, поддерживаемых абонентским радиоинтерфейсом, в попытке найти все точки радиодоступа с требуемыми клиенту идентификатором SSID и поддерживаемыми скоростями радиообмена.

Каждая точка радиодоступа из находящихся в зоне радиовидимости абонента, удовлетворяющая запрашиваемым во фрейме probe request параметрам, отвечает фреймом probe response, содержащим синхронизирующую информацию и данные о текущей загрузке точки радиодоступа. Абонент определяет, с какой точкой радиодоступа он будет работать, путем сопоставления поддерживаемых ими скоростей радиообмена и загрузки. После того, как предпочтительная точка радиодоступа определена, абонент переходит в фазу аутентификации.

Открытая аутентификация (Open Authentication)

Открытая аутентификация по сути не является алгоритмом аутентификации в привычном понимании. Точка радиодоступа удовлетворит любой запрос открытой аутентификации. На первый взгляд, использование этого алгоритма может показаться бессмысленным, однако следует учитывать, что разработанные в 1997 г. методы аутентификации IEEE 802.11 ориентированы на быстрое логическое подключение к беспроводной ЛВС. Кроме того, многие IEEE 802.11-совместимые устройства представляют собой портативные блоки сбора информации (сканеры штрих-кодов и т.п.), не имеющие достаточной процессорной мощности, требующейся для реализации сложных алгоритмов аутентификации [5].

В процессе открытой аутентификации происходит обмен сообщениями двух типов:

- запрос аутентификации (authentication request);
- подтверждение аутентификации (authentication response).



Рис. 2. Открытая аутентификация с несовпадающими WEP-ключами

Таким образом, при открытой аутентификации возможен доступ любого абонента к беспроводной ЛВС. Если в беспроводной ЛВС не используется шифрование, то любой абонент, знающий идентификатор SSID точки радиодоступа, получит доступ к сети. При использовании точками радиодоступа шифрования WEP сами ключи шифрования становятся средством контроля доступа. Если абонент не располагает корректным WEP-ключом, то даже в случае успешной аутентификации он не сможет ни передавать данные через точку радиодоступа, ни расшифровывать данные, переданные точкой радиодоступа (рис. 2).

Аутентификация с общим ключом (Shared Key Authentication)

Аутентификация с общим ключом является вторым методом аутентификации стандарта IEEE 802.11. Аутентификация с общим ключом требует настройки у абонента статического ключа шифрования WEP. Процесс аутентификации иллюстрирует рис. 3.

1. Абонент посылает точке радиодоступа запрос аутентификации, указывая при этом необходимость использования режима аутентификации с общим ключом.
2. Точка радиодоступа посылает подтверждение аутентификации, содержащее challenge text.
3. Абонент шифрует challenge text своим статическим WEP-ключом и посылает точке радиодоступа запрос ассоциирования.
4. Если точка радиодоступа в состоянии успешно расшифровать запрос аутентификации и содержащийся в нем challenge text, она посылает абоненту подтверждение аутентификации, таким образом предоставляя доступ к сети.

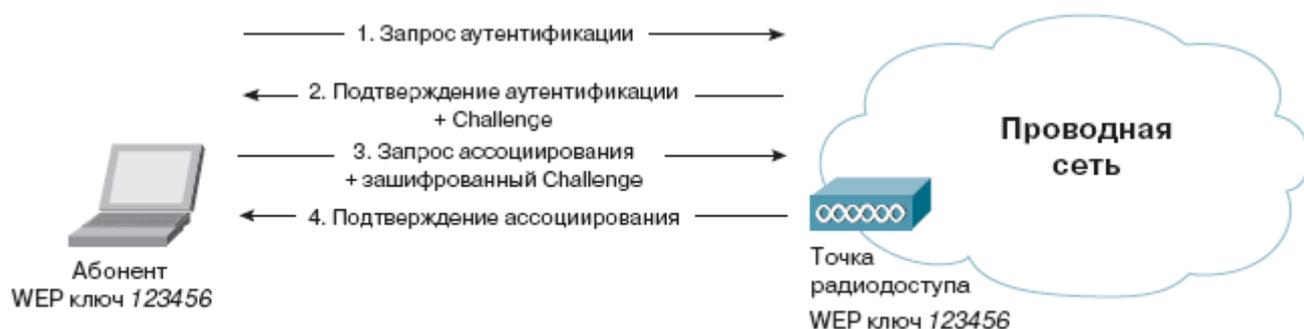


Рис. 3. Аутентификация с общим ключом

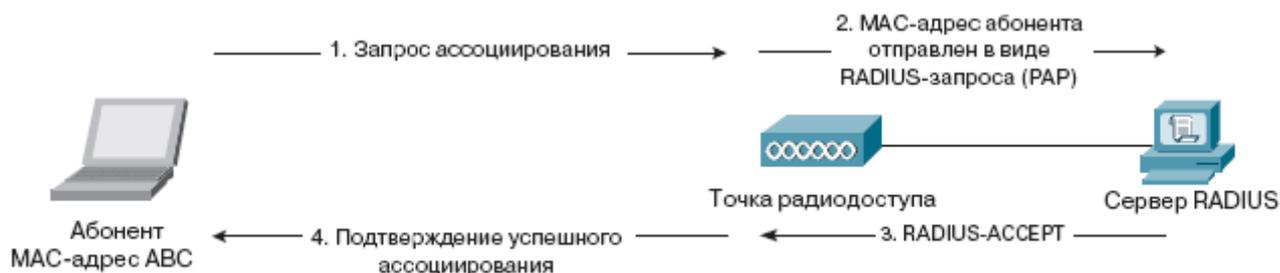


Рис. 4. Аутентификация по MAC-адресу

Аутентификация по MAC-адресу (MAC Address Authentication)

Аутентификация абонента по его MAC-адресу не предусмотрена стандартом IEEE 802.11, однако поддерживается многими производителями оборудования для беспроводных ЛВС, в том числе Cisco Systems. При аутентификации по MAC-адресу происходит сравнение MAC-адреса абонента либо с хранящимся локально списком разрешенных адресов легитимных абонентов, либо с помощью внешнего сервера аутентификации (рис. 4). Аутентификация по MAC-адресу используется в дополнение к открытой аутентификации и аутентификации с общим ключом стандарта IEEE 802.11 для уменьшения вероятности доступа посторонних абонентов.

2. Уязвимость механизмов аутентификации

Проблемы идентификатора беспроводной ЛВС

Идентификатор SSID регулярно передается точками радиодоступа во фреймах beacon. Несмотря на то, что фреймы beacon играют чисто информационную роль в радиосети, т.е. совершенно «прозрачны» для абонента, сторонний наблюдатель в состоянии с легкостью определить SSID с помощью анализатора трафика протокола 802.11, например Sniffer Pro Wireless. Некоторые точки радиодоступа, в том числе Cisco Aironet, позволяют административно запретить широковещательную передачу SSID внутри фреймов beacon. Однако и в этом случае SSID можно легко определить путем захвата фреймов probe response, посылаемых точками радиодоступа.

Идентификатор SSID не разрабатывался для использования в качестве механизма обеспечения безопасности. Вдобавок к этому отключение широковещательной передачи SSID точками радиодоступа может серьезно отразиться на совместности оборудования беспроводных ЛВС различных производителей при использовании в одной радиосети. Вследствие этого Cisco не рекомендует использование SSID в целях реализации режима безопасности [2].

Уязвимость открытой аутентификации

Открытая аутентификация не позволяет точке радиодоступа определить, является ли абонент легитимным или нет. Это становится серьезной брешью в системе безопасности в том случае, если в беспроводной ЛВС не используется шифрование WEP. Cisco не рекомендует эксплуатацию беспроводных ЛВС без шифрования WEP. В случаях, когда использование шифрования WEP не требуется или невозможно (например, в беспроводных ЛВС публичного доступа), методы аутентификации более высокого уровня могут быть реализованы посредством Cisco Service Selection Gateway (SSG) [1].

Уязвимость аутентификации с общим ключом

Аутентификация с общим ключом требует настройки у абонента статического WEP-ключа для шифрования challenge text, отправленного точкой радиодоступа. Точка радиодоступа аутентифицирует абонента посредством дешифрования его ответа на challenge и сравнения его с отправленным оригиналом.

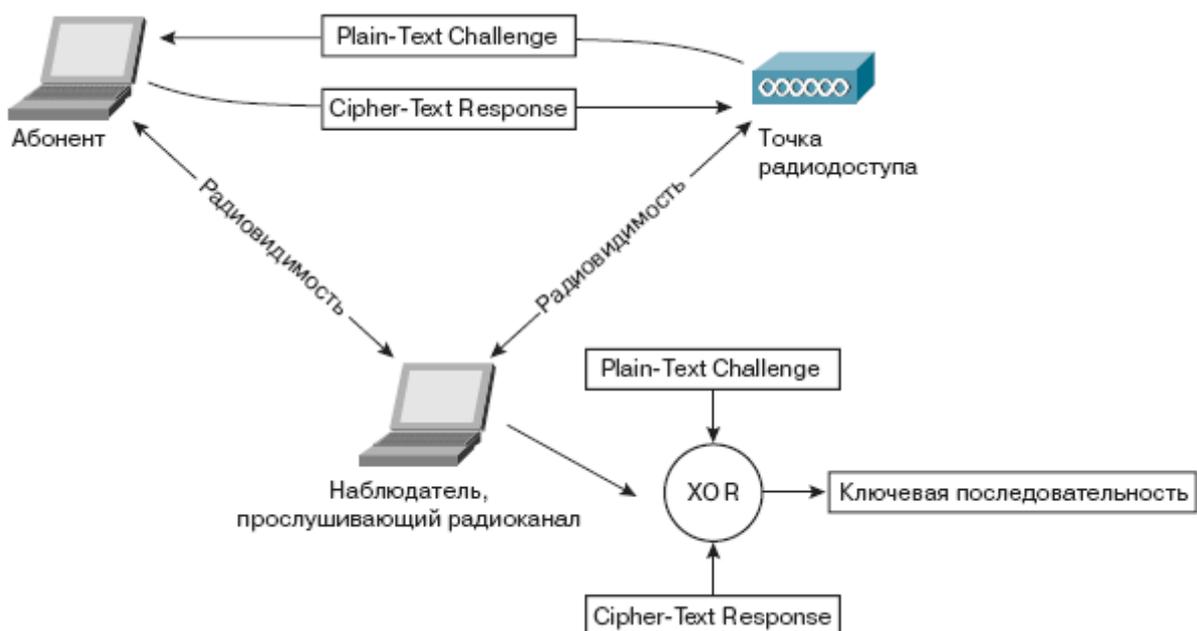


Рис. 5. Уязвимость аутентификации с общим ключом

Обмен фреймами, содержащими challenge text, происходит по открытому радиоканалу и, следовательно, подвержен атакам со стороны стороннего наблюдателя (man-in-the-middle attack). Наблюдатель может принять как нешифрованный challenge text, так и тот же challenge text, но уже в зашифрованном виде (рис. 5). Шифрование WEP производится путем выполнения побитовой операции XOR над текстом сообщения и ключевой последовательностью (key stream), в результате чего получается зашифрованное сообщение (ciphertext). Важно понимать, что выполнение побитовой операции XOR над зашифрованным сообщением и ключевой последовательностью имеет результатом текст исходного сообщения. Таким образом, наблюдатель может легко вычислить сегмент ключевой последовательности путем анализа фреймов в процессе аутентификации абонента.

Уязвимость аутентификации по MAC-адресу

Стандарт IEEE 802.11 требует передачи MAC-адресов абонента и точки радиодоступа в открытом виде. В результате этого в беспроводной ЛВС, использующей аутентификацию по MAC-адресу, хакер может обмануть метод аутентификации путем подмены своего MAC-адреса на легитимный. Подмена MAC-адреса возможна в беспроводных адаптерах, допускающих использование локально администрируемых MAC-адресов. Хакер может воспользоваться анализатором трафика протокола IEEE 802.11 для выявления MAC-адресов легитимных абонентов [4].

Атаки с фрагментацией

Существует достаточно много методов формирования пакетов WEP без знания ключа шифрования. Наиболее эффективным является использование фрагментации на канальном уровне 802.11. Суть этого метода заключается в эксплуатации атаки с известным открытым текстом. Используя предсказуемый формат заголовков LLC, можно восстановить 8 байт гаммы (выхода алгоритма PRGA в RC4, далее PRGA). Для этого первые 8 из зашифрованных байт складываются по модулю 2 с константой, содержащей стандартное значение заголовков LLC (рис. 6).

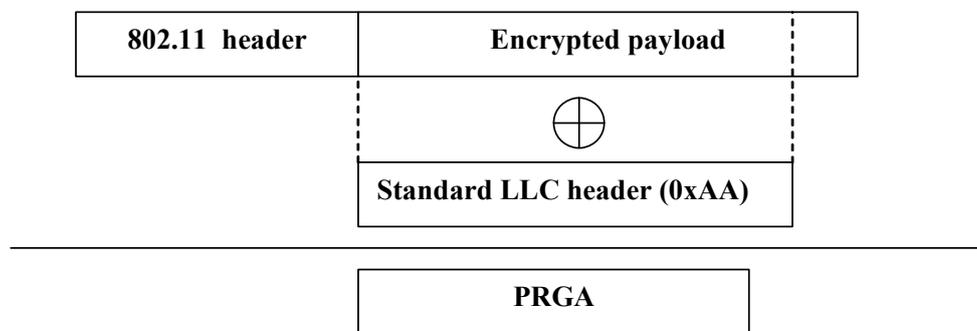


Рис. 6. Восстановление отрезка гаммы

Как видно из рис. 7, в заголовке LLC два последних байта могут меняться. Их значение определяет тип используемого протокола вышестоящего уровня. Возможные значения данных полей описаны в документах IANA.

0xAA	0xAA	0x03	0x00	0x00	0x00	????	????
DSAP	SSAP	CTRL	OU			Ether type	

Рис. 7. Стандартный заголовок LLC

В большинстве случаев беспроводные сети используются для передачи IP-трафика. Следовательно, поле Ether type может принимать одно из трех возможных значений:

- 0x0800 – при передаче IP пакетов;
- 0x0806 – для пакетов ARP;
- 0x86DD – для пакетов IPv6.

Пакеты ARP легко отличить от других по их фиксированному размеру (28 байт данных). Использование протокола IPv6 достаточно просто идентифицируется по наличию широковещательных пакетов на MAC-адреса 33:33:xx:xx:xx:xx, используемых протоколом IPv6 NDP. Полученные 8 байт гаммы могут быть использованы для передачи в сеть произвольных данных той же длины. Но с практической точки зрения это не представляет большого интереса, поскольку все 8 байт в передаваемом пакете будет занимать заголовок LLC. Чтобы обойти это ограничение, может использоваться функция фрагментации на канальном уровне. Беспроводные сети реализуют механизм, позволяющий передать один пакет вышестоящего уровня в нескольких (до 16) фрагментах 802.11. После перехвата одного из пакетов клиента и восстановления PRGA отправляемый пакет разделяется на несколько фрагментов, содержащих по 4 байта данных. Каждый из них передается как отдельный фрейм с использованием функции фрагментации 802.11. Пакеты дополняются контрольной суммой (WEP ICV) и зашифровываются с использованием отрезка восстановленной гаммы. Таким образом, без знания ключа WEP в сеть можно передать пакеты длиной до 64 байт [3].

На практике в сеть можно передать пакеты большего размера. Для этого используется IP-фрагментация, а также структура некоторых служебных пакетов. Например, при перехвате пакета ARP пакета можно восстановить не 8, а 24 байта гаммы. Для этого используются крайне предсказуемые значения заголовков LLC, ARP, а также MAC-адрес отправителя, указанный в заголовках 802.11 в открытом виде (рис. 8).

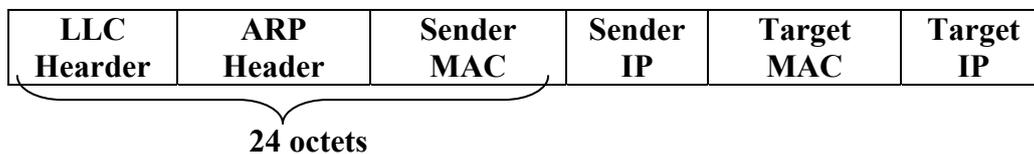


Рис. 8. Пакет ARP

При использовании в сети IPv6 можно восстановить и больший отрезок гаммы. Например, при перехвате пакетов IPv6 NDP Neighbor Solicitation или Router Solicitation можно восстановить до 50 байт гаммы (заголовки LLC + заголовки IP + 2 байта заголовков ICMP). Это связано с тем, что в заголовке IPv6 отсутствует поле контроля целостности. Кроме того, при использовании Local-Link адресации адрес IPv6 можно восстановить по MAC-адресу в заголовках 802.11, если узлом не используются механизмы рандомизации адресов. Отличить разные типы сообщений IPv6 можно по MAC-адресам получателей и размеру. Например, пакет IPv6 Router Solicitation имеет длину 70 байт и передается на MAC-адрес 33:33:00:00:00:02. С использованием 50 байт PRGA в сеть можно передать пакеты размером до 736 байт $((50-4) \times 16)$, что более чем достаточно для практических целей. Генерация трафика при использовании атаки с фрагментацией у злоумышленника, установившего ложную точку доступа, позволяет передать подключившейся станции зашифрованный пакет, которой будет гарантированно обработан получателем. Таким образом, остается только сформировать пакет, на который клиент ответит. Примером подобного пакета является ARP-запрос. Дополнительным плюсом является тот факт, что ARP-пакеты не блокируются персональными межсетевыми экранами. Однако для того, чтобы станция ответила на ARP-запрос, необходимо, чтобы поле Target IP содержало текущий IP-адрес интерфейса. Этой информацией злоумыш-

ленник не обладает, поскольку адрес передается в пакетах в зашифрованном виде. Чтобы получить IP-адрес станции, можно воспользоваться ARP-сканированием, т.е. отправкой ARP-запросов на различные адреса получателей и ожиданием ответа на один из них. Если ответ был получен, значит, станция использует запрошенный IP-адрес.

В качестве диапазонов для сканирования могут выбираться адреса из диапазона APIPA (169.254/32) или распространенные адреса RFC 1918 (например, 192.168.0/24). После того, как IP-адрес станции был определен, используется повторная передача ARP-запроса с целью получения необходимого для KoreK-атак количества пакетов с различными векторами инициализации. В случае поддержки станцией IPv6 ситуация упрощается. Поскольку большинство реализаций стека IPv6 отвечает на широковещательные (например, направленные на адрес ff02::01) ICMPv6-echo запросы, то злоумышленнику достаточно отправить подобный пакет в сеть [1].

Также в IPv6 может применяться пакет IPv6 Neighbor Solicitation (аналог ARP-Request в IPv4). В этом случае осуществлять подбор IP-адреса нет необходимости, поскольку Local-Link IP-адрес может быть определен по MAC-адресу станции.

Уязвимости стандарта беспроводной безопасности 802.11i

Данный стандарт зиждется на двух китах. Первый – уже упомянутый протокол контроля доступа на базе портов 802.1x с надстройками в виде EAP. Использование 802.1x/EAP в основанных на 802.11i протоколах сертификатов беспроводной безопасности WPAv1 и WPAv2 принципиально не отличается. Главное отличие между WPAv1 и WPAv2 – использование принципиально различных систем симметричного шифрования (TKIP в WPAv1, CCMP в WPAv2) и хеширования (MIC в WPAv1, CBC-MAC в WPAv2).

Атаки, не связанные с уязвимостями 802.1x/EAP

В настоящее время WPAv2 является новоиспеченным стандартом, и атаки на WPAv2, не связанные с уязвимостями 802.1x/EAP, остаются теоретическими. Пока они сводятся всего лишь к DoS, например, путем истощения ресурсов беспроводного шлюза с поддержкой WPAv2 при помощи создания множества процессов аутентификации с использованием произвольных MAC-адресов несуществующих клиентов. Еще один вектор DoS атаки против WPAv2 – инъекция подделанного первого пакета четырехпакетного обмена при установлении ассоциации точка доступа – клиент с использованием WPAv2. Эта атака возможна потому, что данный пакет не использует хеширования для проверки целостности пакета во избежание потенциальных атак повтора пакета, если используется общий статический ключ CCMP. Так как известных практических имплементаций и доказательств эффективности этих атак пока не существует, мы не будем заниматься их подробным рассмотрением в этой статье.

WPAv1 является временным решением при переходе от WEP к WPAv2, не требующем модернизации аппаратной части. Помимо теоретических, не связанных с 802.1x, атак против WPAv1, таких как атака на хэш временного ключа (снижающая сложность извлечения ключа с 2^{128} до 2^{105}) и DoS атака искажения контрольной суммы MIC (реализация которых намного сложнее, чем кажется на первый взгляд), существуют и прикладные атаки против WPAv1-SOHO, использующего предварительно разделенный ключ, общий для всех узлов с одним ESSID в большинстве реализаций WPAv1-SOHO. Первая атака представляет собой генерацию временных ключей других клиентских узлов, если известен постоянный общий ключ (PSK). Таким образом, практически эта атака представляет ценность для легитимного пользователя сети, который желает прослушивать и манипулировать трафиком других легитимных пользователей (вариант сотрудника, атакующего соединение своего руководителя). Несмотря на то, что каждый узел на сети, защищенной WPAv1-SOHO, имеет свой зашифрованный ка-

нал соединения с точкой доступа, временные ключи для защиты этого канала генерируются с помощью PSK, двух случайных величин из двух первых пакетов четырехстороннего квитирования WPAv1-SOHO и MAC адресов участвующих узлов. Таким образом, атакующий, уже обладающий PSK, может легко перехватить MAC-адреса вовлеченных узлов и инициировать процедуру квитирования с помощью DoS атаки фреймами деассоциации для перехвата первых двух пакетов обмена с нужными величинами. Имея эти данные под рукой, несложно сгенерировать временный ключ для атакуемого канала.

Если же атакующий не знает PSK, он может воспользоваться атакой перебора по словарю или даже случайного перебора против временного ключа, а затем, имея величины, упомянутые выше, сгенерировать PSK из угаданного временного ключа, осуществив действия первой описанной атаки в обратном порядке.

Атаки против 802.1x/EAP

Данные атаки можно подразделить на атаки против 802.1x, вне зависимости от используемого типа EAP, и атаки против отдельных EAP разновидностей. К первым относится посылка фальшивых EAP-Failure и EAPOL-logoff (EAP через локальную сеть) фреймов, затопление фреймами EAPOL-Start и циклическим перебором идентификаторов EAP, а также преждевременной отправкой фреймов EAP-Success. Так как принцип работы подобных атак понятен, и мы не особенно заинтересованы в DoS атаках (хотя они могут иметь большое значение в проведении атак «человек в середине», всегда есть «старые добрые» фреймы деаутентификации), мы не будем акцентироваться на них и перейдем к атакам на специфические типы EAP. Самый первый стандартизированный тип EAP – это EAP-MD5, который использует схему аутентификации, аналогичную аутентификации CHAP. Сейчас EAP-MD5 практически вышел из употребления и может быть встречен, в основном, в случае режима подстраховки, когда по какой-то причине более совершенные типы EAP не работают. Основной уязвимостью EAP-MD5 является отсутствие какой-либо аутентификации с «серверной» стороны, сопряженное с отсутствием туннелирования трафика этого протокола. Таким образом, кракер может представить свою «пиратскую» точку доступа с большей силой сигнала и сопряженным RADIUS сервером, «переманив» клиентские машины на ее сторону после массовой DoS атаки фреймами деаутентификации. Подобные атаки на MD4 хэши длиной всего 6 байт отличаются значительной скоростью, позволяющей использовать радужные таблицы больших размеров. В настоящее время Cisco рекомендует использовать более новый и безопасный EAP-FAST вместо EAP-LEAP.

Заключение

На настоящий момент относительно безопасными можно считать только сети стандарта 802.11, защищенные с помощью WPAv1 с 802.1x и WPAv2 с 802.1x при условии использования типов EAP с поддержкой туннелирования и взаимной аутентификацией обоих концов туннеля. К таким типам EAP относятся EAP-TLS и EAP-FAST. При этом EAP-TLS требует наличия сертификатов аутентификации на всех клиентских хостах, что делает установку и менеджмент массивных сетей, защищенных с использованием этого протокола, весьма трудоемкой. В то же время EAP-FAST поддерживается по преимуществу аппаратным обеспечением Cisco и требует покупки дополнительного программного обеспечения для поддержки систем, отличных от Windows XP, Windows 2000 и Windows CE. Таким образом, дизайн защищенных сетей стандарта 802.11 является более сложной задачей, чем представляют себе многие архитекторы и администраторы таких сетей, даже при рассмотрении исключительно протоколов 802.11i и без учета альтернатив, таких как IPSec.

Литература

1. The Radical Realm of RADIUS, 802.1x, and You. 2007. Rodney Thayer, Beetle, Shmoo Group, LayerOne.
2. Message Attack on the 4-Way Handshake. 2006. ChangHua He, John C. Mitchell. Stanford University.
3. Attacks against Michael and Their Countermeasures. 2004. Dan Harkins. Trapeze Networks.
4. Weakness in a Temporal Key Hash of WPA. 2006. Vebjorn Moen at al., Bergen University.
5. Fast and Secure Roaming in WLAN. 2007. Magnus Falk, Linkoping University.

ВОЗМОЖНЫЕ КАНАЛЫ УТЕЧКИ ИНФОРМАЦИИ НА ПРЕДПРИЯТИИ

А.В. Годырева

Научный руководитель – к.т.н., доцент Н.С. Кармановский

Для сохранения информации следует определить потенциальные и реальные каналы утечки информации. Утечка информации может происходить как из-за неорганизованности и небрежности в работе сотрудников предприятия, так и в результате экономического и промышленного шпионажа. В данной статье рассматриваются возможные каналы утечки информации и меры по предотвращению утечки конфиденциальной информации.

Известно, что до 70–80% утечки информации связаны с действием сотрудников компании, работающих на предприятии, а также уволенными сотрудниками. Практика показывает, что мотивация совершения подобных действий может быть самой различной. С повышением значимости и ценности информации растет и важность ее защиты. Утечка или утрата информации влечет за собой материальный ущерб. Кроме того, информация – это фактор управления. Несанкционированное вмешательство в управление может привести к катастрофическим последствиям в объекте управления – производстве, транспорте, военном деле. Если в организации присутствует конфиденциальная информация, то в ней, в первую очередь, должны быть приняты меры по защите информации от различных угроз.

Ущерб от раскрытия конфиденциальной информации может выражаться в потере преимуществ, упущенной коммерческой выгоде, санкциях со стороны органов управления, административной и уголовной ответственности, ухудшении морального климата в коллективе вследствие раскрытия информации о заработной плате работников, о планируемых кадровых перестановках и т.п.

Конфиденциальная информация – это информация, доступ к которой ограничен кругом доверенных пользователей.

Коммерческая тайна – конфиденциальность информации, позволяющая ее обладателю при существующих или возможных обстоятельствах увеличить доходы, избежать неоправданных расходов, сохранить положение на рынке товаров, работ, услуг или получить иную коммерческую выгоду [1].

Защита от утечки конфиденциальной информации – задача достаточно сложная. Возрастающее использование электронной почты, интернет-пейджеров и других средств передачи данных, распространенность мобильных устройств, с помощью которых сотрудники могут выносить важную информацию за пределы организации – все это значительно осложняет контроль над потоками данных. Конкретное определение угрозы и стратегия ее предотвращения – обусловлены тремя основными вопросами.

- Какая информация является конфиденциальной и нуждается в защите от несанкционированного разглашения?
- Как может произойти утечка информации, т.е. где она хранится, в каких производственных процессах используется, какие точки выхода следует защитить?
- Каким группам пользователей разрешен доступ к конфиденциальной информации, какие пользователи являются наиболее вероятными «кандидатами», способными создать брешу для утечки информации, и нуждаются в более пристальном наблюдении [2]?

Виды угроз

Угрозы можно разделить на два класса: естественные (объективные) и искусственные (субъективные). Естественные угрозы – это угрозы, не связанные с деятельно-

стью человека. Искусственные угрозы – это угрозы, вызванные деятельностью человека. Искусственные угрозы можно разделить на:

- непреднамеренные угрозы – несознательно наносимый вред. Субъект таких угроз называется нарушителем;
- преднамеренные или умышленные угрозы – это угрозы, при которых человек сознательно совершает вред. Субъект таких угроз называется злоумышленником.

Умышленные угрозы могут проявляться различными путями:

- внедрение злоумышленника в персонал предприятия;
- вербовка персонала;
- несанкционированное проникновение на территорию объекта;
- дистанционное воздействие и наблюдение.

В таблице содержится информация об угрозах, которые может совершить нарушитель и злоумышленник [3].

Виды угроз	Субъекты угроз			
	Стихия	Нарушитель	Злоумышленник	
			На территории предприятия	Вне территории предприятия
Травмы и гибель людей	+	+	+	+
Повреждение оборудования и техники	+	+	+	+
Повреждение системы жизнеобеспечения	+	+	+	+
Несанкционированное изменение технологии производства		+	+	
Использование нерегламентированных технических и программных средств		+	+	
Хищение материальных ценностей			+	
Уничтожение или перехват данных путем хищения носителей			+	
Устное разглашение конфиденциальной информации		+		
Несанкционированный съем информации			+	+
Нарушение правил эксплуатации средств защиты		+	+	

Таблица. Виды угроз

Наиболее опасные виды информационных угроз – это несанкционированный сбор информации, модификация информации, утечка и уничтожение.

Разделим все источники угроз на внешние и внутренние. Источниками внутренних угроз являются:

- сотрудники организации;
- программное обеспечение;
- аппаратные средства.

Внутренние угрозы могут проявляться в следующих формах:

- ошибки пользователей и системных администраторов;
- нарушения сотрудниками фирмы установленных регламентов сбора, обработки, передачи и уничтожения информации;
- ошибки в работе программного обеспечения;
- отказы и сбои в работе компьютерного оборудования.

К внешним источникам угроз относятся:

- компьютерные вирусы и вредоносные программы;
- деятельность отдельных «внешних» организаций и связанных с ними лиц;
- стихийные бедствия.

Формами проявления внешних угроз являются:

- заражение компьютеров вирусами или вредоносными программами;
- несанкционированный доступ (НСД) к корпоративной информации;
- информационный мониторинг со стороны конкурирующих структур, разведывательных и специальных служб;
- действия государственных структур и служб, сопровождающиеся сбором, модификацией, изъятием и уничтожением информации;
- аварии, пожары, техногенные катастрофы.

Отдельно следует рассмотреть информационную угрозу, связанную с деятельностью конкурентов. Угроза может проявляться в виде переманивания специалистов, внедрения своих специалистов на предприятие, проведения дистанционного съема информации. Реальную угрозу организации представляют уволенные сотрудники. Уволенные сотрудники могут передавать информацию конкурентам, создавать аналогичное конкурентное предприятие. Одним из важнейших направлений профилактики утечки информации является своевременное заключение с сотрудниками предприятия соглашений о неразглашении конфиденциальной информации, что защищает фирму юридически.

Каналы утечки информации

Наиболее распространенные из них имеют характер неумышленного раскрытия информации сотрудниками по причине слабой профилактической работы на предприятии, слабого владения правилами работы с конфиденциальной информацией и неумением определять степень ее конфиденциальности.

Каналы утечки информации можно разбить на четыре группы [3].

Первая группа – каналы, связанные с доступом к элементам системы обработки данных, но не требующие изменения компонентов системы. Это каналы, образующиеся за счет перехвата электромагнитных излучений и наводок, применения подслушивающих устройств и т.п.

Вторая группа – каналы, связанные с доступом к элементам системы и изменением структуры ее компонентов. К ней относятся: хищение и копирование носителей информации; наблюдение за информацией в процессе обработки с целью ее запоминания; чтение остаточной информации, т.е. данных, остающихся на магнитных носителях после выполнения заданий.

Третья группа – незаконное подключение специальной регистрирующей аппаратуры к устройствам системы или линиям связи; злоумышленный вывод из строя меха-

низмов защиты; злоумышленное изменение программ таким образом, чтобы эти программы наряду с основными функциями обработки информации осуществляли также несанкционированный сбор и регистрацию защищаемой информации.

Четвертая группа – несанкционированное получение информации путем подкупа или шантажа должностных лиц соответствующих служб; получение информации путем подкупа и шантажа сотрудников, знакомых, обслуживающего персонала или родственников, знающих о роде деятельности.

Представленная классификация каналов утечки конфиденциальной информации показывает очевидность того, что проблему утечки информации нельзя решить каким-либо одним способом, тем более решить ее полностью. Кроме того, реализация любых мер по ограничению доступа к информации или ее распространению потенциально снижает эффективность основных бизнес-процессов организации. Оптимальный подход заключается в сочетании разумного доступа со строгой регламентацией работы с конфиденциальной информацией. В данном случае это означает необходимость создания системы организационно-технических мероприятий, позволяющих перекрыть основные каналы утечки информации с определенной степенью надежности и минимизировать существующие риски без существенного снижения эффективности бизнес-процессов. В отсутствие такой системы информационные риски для бизнеса неприемлемо высоки, а права на юридическую защиту интересов организации, как собственника информации, нереализуемы.

Способы предотвращения утечки конфиденциальной информации

Система предотвращения утечки конфиденциальной информации может включать в себя три основные составляющие.

- **Работа с персоналом.**

Основным источником утечки информации является персонал. Человеческий фактор способен свести на нет любые самые изощренные механизмы безопасности. Основные принципы и правила управления персоналом определены в международном стандарте ISO/IEC 17799:2000. Соблюдение этих правил позволяет существенно снизить влияние человеческого фактора, избежать характерных ошибок и предотвратить утечку информации [2].

- **Политика безопасности.**

Политика безопасности организации – совокупность руководящих принципов, правил, процедур и практических приемов в области безопасности, которыми руководствуется организация в своей деятельности.

Основные направления разработки политики безопасности:

- классификация данных по необходимой степени защищенности;
- определение субъекта и размера ущерба, наносимого фирме в информационном аспекте,
- вычисление рисков и определение схемы уменьшения их до приемлемой величины.

- **Сервисы безопасности.**

Сервисы безопасности используются для ограничения доступа к информации, протоколирования фактов осуществления доступа и контроля информационных потоков. Они позволяют обеспечить предупреждение, предотвращение, обнаружение и реагирование на инциденты, связанные с утечкой информации. К числу сервисов безопасности относятся: сервисы аутентификации, управления доступом, шифрования, и аудита безопасности [3].

Любой руководитель заинтересован в предотвращении утечки конфиденциальной информации. Прежде чем предпринимать меры по защите конфиденциальной информации, следует определить ее ценность и постараться, чтобы затраты на обеспечение ее

конфиденциальности не превосходили ее реальную стоимость. К основным способам предотвращения потери ценной информации относятся:

- организационные;
- правовые;
- технические.

Основные организационные мероприятия по обеспечению предотвращения утечки информации можно условно разделить на два вида:

- определение концепции кадровой политики и критериев подбора персонала; подбор служащих, текущая работа с постоянными сотрудниками, и их социальная защита;
- реализация мер по определению степени конфиденциальных сведений и обеспечение разграничения доступа: классификация документов по степени важности и конфиденциальности, определение режима их хранения, допуска к ним, учета.

Организационные меры и мероприятия связаны с управлением предприятием и относятся к таким действиям, которые сводят к минимуму информационную уязвимость предприятия. К компетенции администрации предприятия относится создание оптимальных условий, обеспечивающих безопасность предприятия.

Меры по созданию оптимального режима, обеспечивающие безопасность предприятия, включают следующие действия:

1. организацию пропускного режима на предприятие;
2. организацию соответствующего режима конфиденциальности в соответствии с законодательством; регистрацию сотрудников, включая посетителей, имеющих доступ к конфиденциальной информации;
3. разграничение документации по степени конфиденциальности и разграничение доступа к информации;
4. регулярные профилактические работы с работниками предприятия;

С этой целью на предприятии желательно создать специальную аналитическую группу информационной безопасности, которая будет:

- заниматься определением ценности имеющейся информации;
- изучать клиентов фирмы;
- определять потенциальных конкурентов;
- осуществлять контроль за рекламой, деловыми контактами, следить за конкуренцией и т.п.

Что касается правового аспекта предотвращения потери ценной информации, следует опираться на нормы действующего законодательства и договорные отношения.

Важной основой правовой защиты коммерческой информации является заключение коммерческих договоров и соглашений, включающих вопросы обеспечения конфиденциальности. Технические средства защиты информации даже при их высокой стоимости не гарантируют полную защиту информации без реализации перечисленных выше мер. К техническим средствам защиты информации относятся разнообразные механические, электромеханические, электронные и другие устройства и системы, которые в совокупности с другими средствами способствуют защите информации предприятия. Технические средства позволяют автоматизировать охрану помещений и снизить возможность утечки информации по другим каналам.

При подборе технических средств защиты следует исходить из разумных соображений и возможностей предпринимателя. Прежде всего, необходимо предусмотреть:

- физическую изоляцию территорий, зданий, помещений;
- оборудование окон решетками, установку специальных дверей;
- оснащение их датчиками систем охранной сигнализации;
- выделение специализированных помещений для техники;
- защиту носителей информации и аппаратуры от похищения;

- регламентацию технологических процессов;
- разработку инструкций для персонала;
- применение специальных технических средств, исключающих прослушивание и перехват информации;
- контроль и проверку средств защиты.

Заключение

Утечка или утрата информации влечет за собой материальный ущерб. В данной статье рассмотрены вопросы, связанные с организацией защиты от утечки конфиденциальной информации. Проведена классификация существующих угроз и каналов утечки информации, использование которых влечет серьезные последствия для предприятия. Существуют каналы утечки информации, образующиеся за счет перехвата электромагнитных излучений и наводок, применения подслушивающих устройств, хищения и копирования носителей информации; наблюдения за информацией в процессе обработки с целью ее запоминания; чтения остаточной информации, незаконного подключения специальной регистрирующей аппаратуры к устройствам системы или линиям связи; злоумышленного вывода из строя механизмов защиты; злоумышленного изменения программ, несанкционированного получения информации путем подкупа или шантажа должностных лиц соответствующих служб; получения информации путем подкупа и шантажа сотрудников, знакомых, обслуживающего персонала или родственников, знающих о роде деятельности.

Определено, что проблему утечки информации нельзя решить каким-либо одним способом. Необходимо создание организационно-технической системы, позволяющей перекрыть каналы утечки конфиденциальной информации. Основными составляющими такой системы являются политика безопасности, работа с персоналом, сервисы безопасности. Рассмотрены способы предотвращения утечки конфиденциальной информации: организационные, правовые и технические мероприятия.

Литература

1. Федеральный закон Российской Федерации от 29 июля 2004 г. №98-ФЗ О коммерческой тайне.
2. Максимов Ю.Н., Сонников В.Г. и др. Шпионские штучки. Технические методы и средства защиты информации. – СПб: Полигон, 2000. – 320 с.
3. Волокитин А.В., Маношкин А.П., Солдатенков А.В., Савченко С.А., Петров Ю.А. Информационная безопасность государственных организаций и коммерческих фирм. – М.: НТЦ «ФИОРД-ИНФО», 2002. – 272с.

СИСТЕМА СКРЫТОЙ ПЕРЕДАЧИ СООБЩЕНИЙ В ГРАФИЧЕСКИХ ИЗОБРАЖЕНИЯХ ФОРМАТА JPEG

С.С. Кувшинов

Научный руководитель – д.т.н., профессор А.Г. Коробейников

В работе предлагается реализация задачи скрытой передачи данных в графических изображениях формата JPEG. Приводится общая схема решения, программная реализация алгоритмов внедрения/извлечения данных и пример работы системы с конкретным сообщением и изображением. Также в работе рассмотрены особенности структуры файлов формата JPEG и особенности межформатных преобразований.

Введение

Наряду с криптографическими, задачу обеспечения конфиденциальности информации решают и стеганографические методы, которые позволяют скрытно передавать данные, т.е. скрывать сам факт передачи сообщений. Интерес к стеганографии появился в последнее десятилетие и вызван широким распространением мультимедийных форматов. В работе предлагается алгоритм встраивания текстовых сообщений в графические изображения формата JPEG. Актуальность данного решения наряду, с существующими стеганографическими системами, оперирующими с файлами BMP, прежде всего, обоснована фактом большего распространения формата изображений JPEG как поддерживающего сжатие. Файлы JPEG (фотографии, графика) распространены в сети Интернет, и их передача является обыденной операцией. JPEG популярен ввиду малого размера файлов данного формата, поэтому обмен JPEG-изображениями не вызывает подозрения факта скрытой передачи информации. Коллекция способов встраивания сообщения в графические стегоконтейнеры многообразна и пополняется с каждым днем новыми модификациями алгоритмов.

Проблематика

В нашей системе сообщение встраивается в графический файл-контейнер, далее файл доставляется адресату, адресат извлекает из полученного изображения текст сообщения. Описанная последовательность действий представлена на рис. 1.

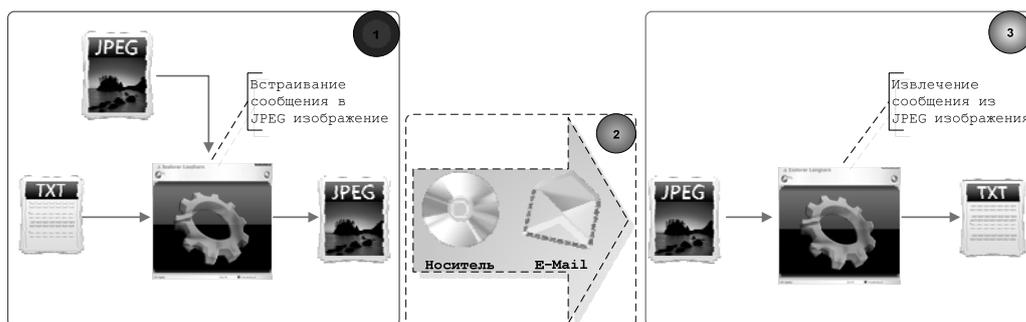


Рис. 1. Общая схема работы системы скрытой передачи

Как показано на рисунке, передача сообщения проходит три этапа.

- 1) Встраивание сообщения в изображение формата JPEG.
- 2) Непосредственно передача полученного изображения с внедренным сообщением адресату (на сменном носителе, по электронной почте).
- 3) Извлечение сообщения из изображения.

Процессы встраивания и извлечения автоматизированы и выполняются с помощью разработанного приложения. Этап 2 является неконтролируемым, и методы из-

влечения и встраивания не зависят от канала передачи. В результате получатель получит извлеченное сообщение, идентичное передаваемому, при условии, что передаваемое изображение с внедренным сообщением (стегопосылка) не менялось с момента завершения процесса встраивания данных отправителем до момента начала их извлечения получателем. При несоблюдении указанной целостности программа не гарантирует тот факт, что получатель сможет извлечь полезную для него информацию (сообщение).

Алгоритм внедрения

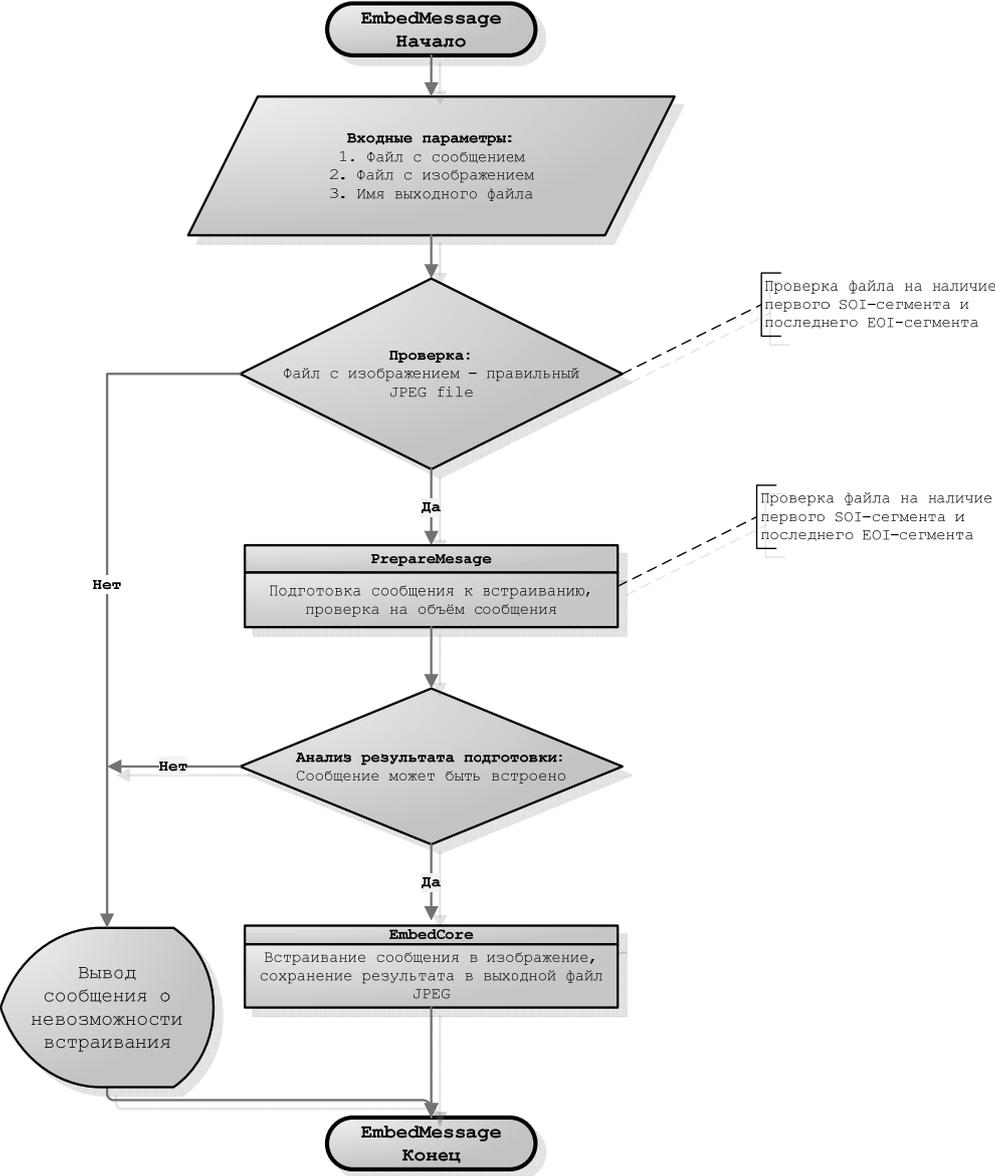


Рис. 2. Общая схема алгоритма внедрения сообщения

Опытным путем определилось, что структура JPEG нерегулярна от файла к файлу. Рассмотренная выше иерархия маркеров присуща в том или ином объеме всем JPEG-файлам, однако каждый отдельно взятый файл имеет свои особенности. К примеру, в каких-то файлах заявлена одна таблица квантования и, соответственно, один блок данных сканирования, в других файлах – несколько таблиц и несколько блоков [1, 2]. В редких случаях встречаются файлы с маркерами перезапуска, что усложняет процедуру

анализа и поэлементного разбора JPEG-структуры. Также файл JPEG может хранить в себе несколько уменьшенных копий изображения для предварительного просмотра; этот факт увеличивает количество сегментов данных изображения, что сказывается на времени анализа файла.

В качестве подготовительной работы были созданы алгоритмы разбора (парсинга) файлов JPEG и BMP и реализованы программы-парсеры. Возможности данных модулей позволяют получить структурированное представление блоков для дальнейшего анализа потенциальных мест для внедрения битов сообщения.

Общая схема алгоритма внедрения представлена на рис. 2

Система принимает входные данные, проверяет их корректность:

1. существование входных файлов на носителе;
2. соотношение размеров входных файлов;
3. соответствие файла-контейнера формату JPEG.

Подготовка сообщения к внедрению

Согласно требованиям к системе, перед внедрением сообщение должно быть зашифровано и сжато. Подготовленное таким образом сообщение должно быть проанализировано на возможность встраивания с учетом требования ограничения объема ($\frac{1}{4}$ от объема изображения-контейнера).

Общая схема подготовки сообщения представлена на рис. 3.

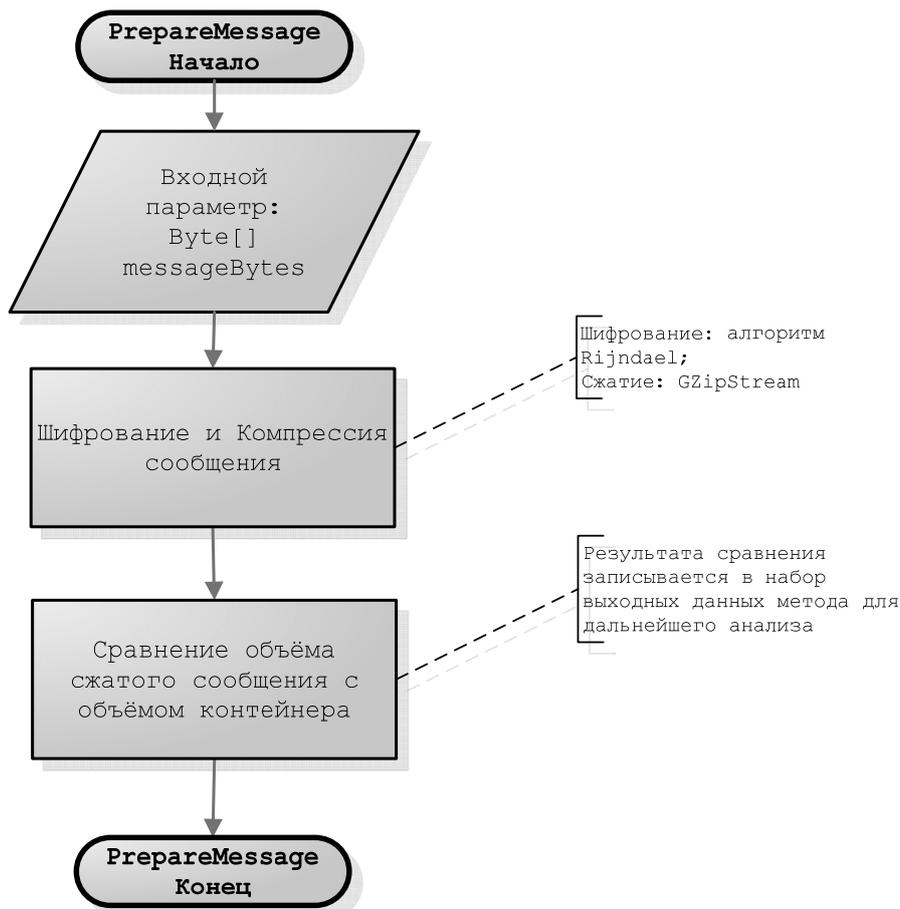


Рис. 3. Алгоритм подготовки сообщения к внедрению

Внедрение сообщения

При разработке алгоритма внедрения обнаружались трудности, касающиеся межформатных преобразований. Внедрение происходит по алгоритму, представленному на рис. 4.

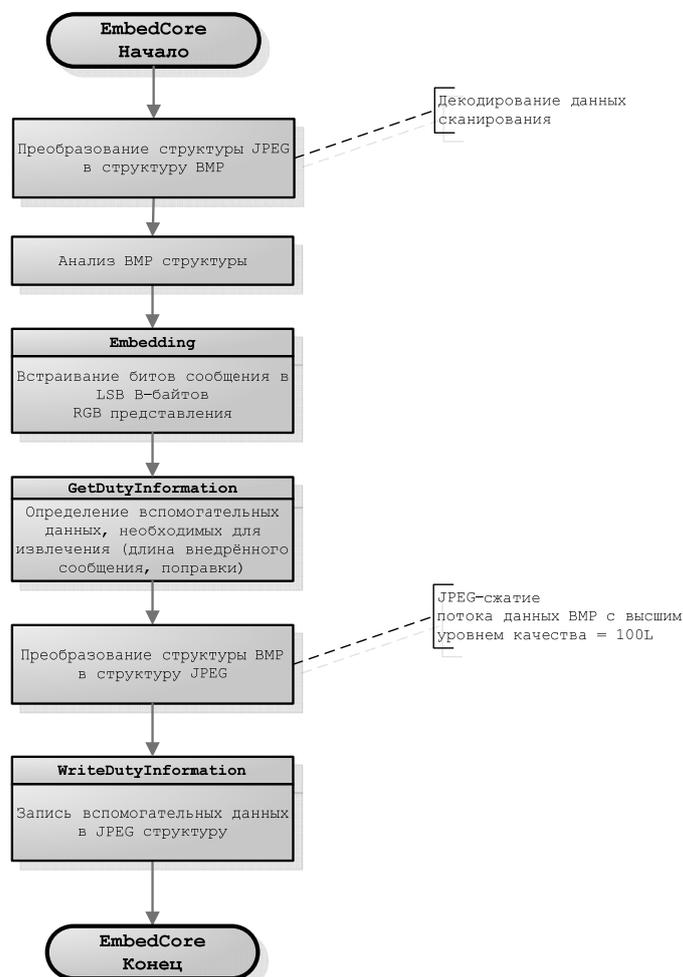


Рис. 4. Порядок действий при внедрении

Сложность разработки алгоритма сокрытия сообщения в файле JPEG заключается в специфичной структуре файла данного формата. Основным объемом в файле данного формата занимают данные сканирования – это информация, в которой закодировано цветное представление изображения. Для просмотра изображения JPEG эти байты должны быть декодированы для восстановления RGB представления изображения. Декодеры, используемые в программах просмотра JPEG изображений, преобразуют данные сканирования с помощью служебной информации, также хранящейся в JPEG файле. Однако самая большая проблема в том, что JPEG – формат сжатия с потерями, и они в общем случае не позволяют восстановить встроенное сообщение, поскольку восстановление происходит после процедур межформатных преобразований JPEG – RGB BMP – JPEG. В разработанной системе указанные проблемы решены.

Суть предлагаемого алгоритма представлена в форме диаграммы перехода состояний изображения на рис. 5.

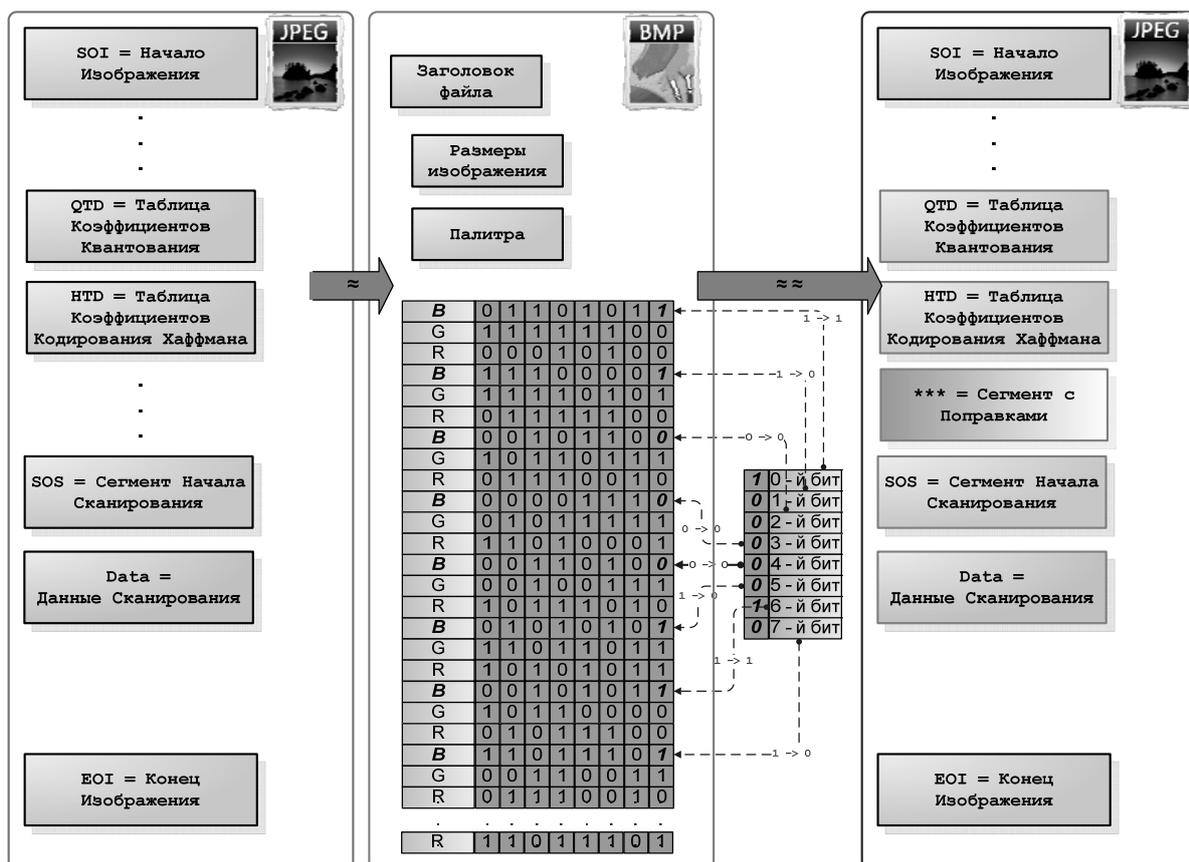


Рис. 5. Последовательность процесса внедрения сообщения

На первом этапе производится преобразование потока данных JPEG в поток данных BMP. При этом увеличивается размер потока за счет изменения принципа кодирования информации об цветовых свойствах участков изображения. За счет того, что в формате BMP каждая точка изображения кодируется тремя байтами [3], отвечающими за вклад основных цветов (R – красного, G – зеленого и B – синего) в целевой цвет точки, изменение размера потока в большую сторону значительно и позволяет встроить необходимый объем информации в себе.

Известно, что человеческий глаз наименее чувствителен к изменениям в оттенках синего цвета, поэтому для встраивания используются B-составляющие RGB структур [4]. Для минимизации объема изменения используется только 1 наименьший бит такого байта, что до минимума снижает вероятность обнаружения изменения даже на изображениях с большой площадью заливки синего цвета. Простейший способ замены битов – последовательная замена в каждом b-байте [3, 5], представлен на рис. 6.

После встраивания выполняется последовательность действий, представленная на рис. 7.

Измененные b-байты выделяются из потока и копируются во временный буфер – буфер 1. Также в другой временный буфер (буфер 2) сохраняется информация о том, как встраивались биты сообщения – в каждый байт подряд или через один, два и т.д. Эта процедура может не выполняться, если порядок встраивания – последовательная замена бита в каждом b-байте.

Далее модифицированный поток байтов BMP подвергается JPEG сжатию с наивысшим коэффициентом качества. Полученный JPEG поток на этой стадии еще не сохраняется в выходной файл. Для обеспечения возможности последующего извлечения производятся дополнительные действия.

Производится попытка извлечения сообщения, которая включает в себе:

1. клонирование потока JPEG во временный поток;
2. декодирование временного потока JPEG в поток байтов BMP;
3. извлечение модифицированных b-байтов, в которые были встроены биты сообщения, с учетом информации из буфера 2.

Далее следует анализ извлеченного сообщения, т.е. сравнение извлеченных b-байтов с b-байтами, сохраненными в буфер 1, и сохранение разностей в новый буфер – буфер 3. Эти разности являются неизбежным результатом потерь при межформатных преобразованиях JPEG – RGB BMP – JPEG.

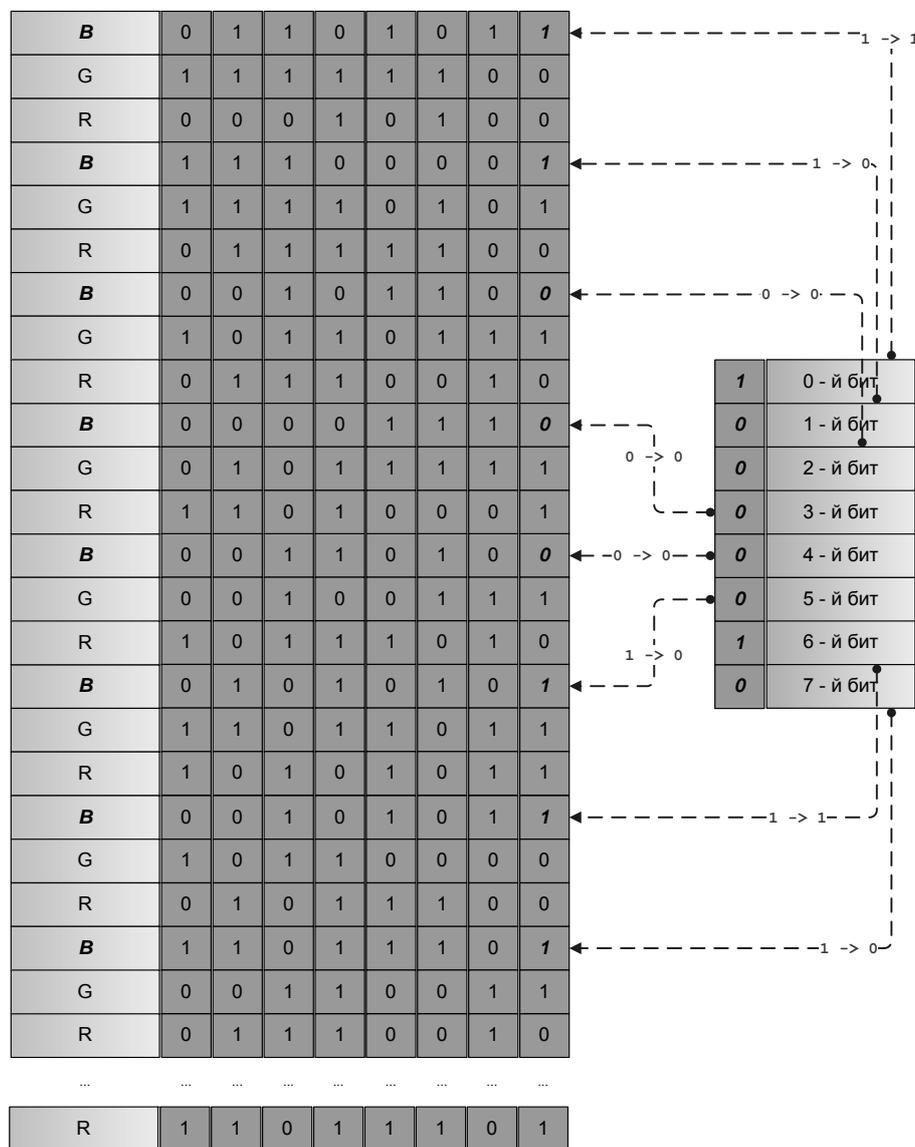


Рис. 6. Последовательная замена битов

Суть подготовительной работы для обеспечения последующего извлечения – это механизм компенсирования потерь. Данные из буфера 3 записываются в СОМ-сегмент JPEG, сегмент добавляется к выходному потоку JPEG, и весь поток сохраняется в целевой файл. Как правило, за счет наивысшего коэффициента качества преобразования из BMP в JPEG значения поправок механизма компенсирования не превышают значения 6 по абсолютной величине, а, следовательно, эта информация представляет собой значения, лежащие в диапазоне от минус шести до плюс шести, включая границы. При использовании усложненного механизма замены битов данные из буфера 2 также сохра-

няются в структуре JPEG в виде нового сегмента и используются программой при извлечении. Важно отметить, что в качестве сегмента, используемого механизмом компенсации потерь, может быть не только сегмент COM. В этой роли может выступать так же любой сегмент, игнорируемый программами просмотра JPEG изображений:

- SOF2 ÷ SOF10;
- DAC;
- DNL;
- APP15;
- Пользовательские (неспецифицированные) типы сегментов [1].

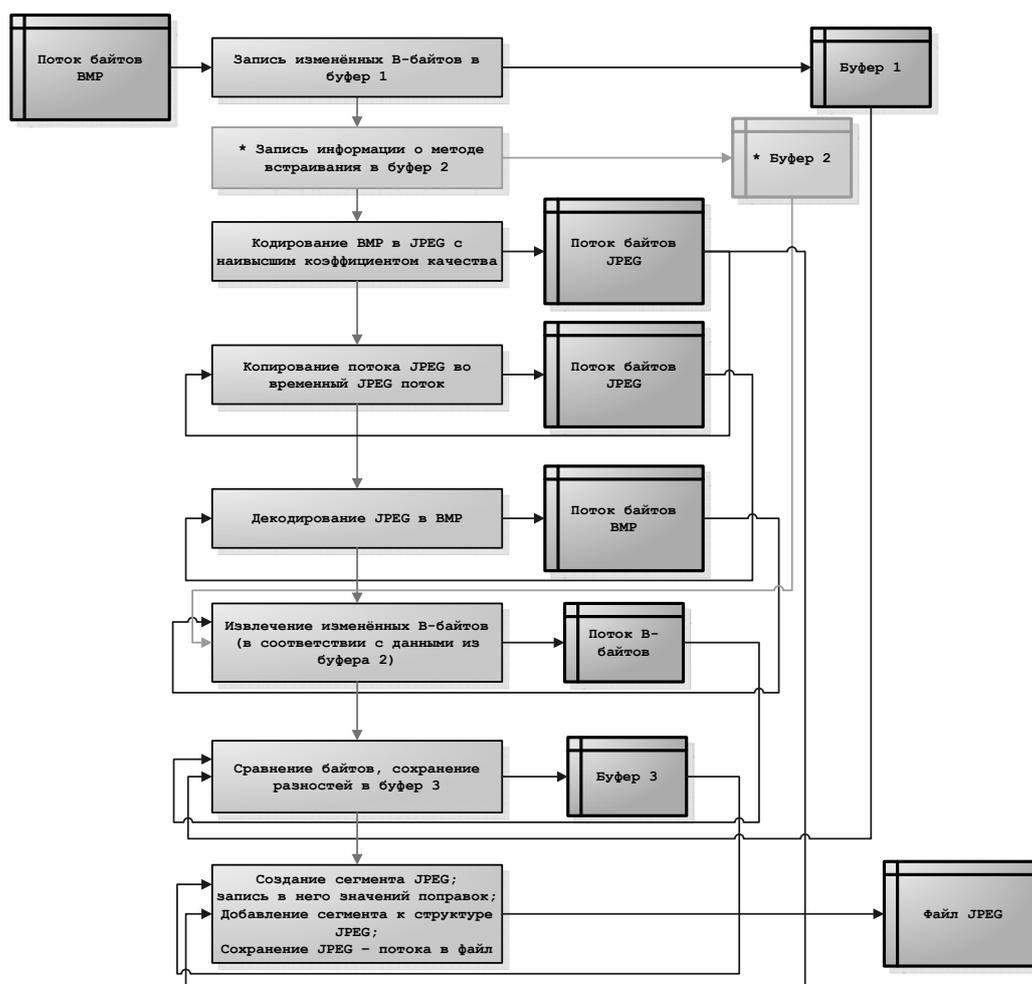


Рис. 7. Реализация механизма компенсации потерь (расчет и сохранение байтов-поправок)

Заключение

В ходе исследований и тестирования системы получены результаты, удовлетворяющие требованиям технического задания:

1. шифрование и сжатие сообщения перед внедрением в контейнер;
2. визуальные изменения изображения с внедренным сообщением относительно исходного изображения должны быть незаметны человеческому глазу;
3. работа с сообщениями, размер которых при внедрении составляет до $\frac{1}{4}$ размера изображения-контейнера.

Литература

1. Code net, всё для программиста [Электронный ресурс] / JPEG – Алгоритм сжатия Web-мастер: О.С. Савченко, 2001. – Режим доступа: <http://www.codenet.ru/progr/formt/intro.php>, свободный. – Загл. с экрана. – Яз. рус.
2. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. – М.: ДИАЛОГ-МИФИ, 2003. – 348 с.
3. Омский государственный университет в дополнительных страницах [Электронный ресурс] / Омский государственный университет, Web- мастер: О.С. Савченко, 2001. – Режим доступа: <http://www.univer.omsk.su/omsk/Edu/infpro/1/jpeg/fbmp.htm>, свободный. – Загл. с экрана. – Яз. рус.
4. Компьютерра Онлайн [Электронный ресурс] / Особенности цветового восприятия человека, 2001. – Режим доступа: <http://offline.computerra.ru/1998/273/1965/>, свободный. – Загл. с экрана. – Яз. рус.
5. Компьютерная документация от А до Я [Электронный ресурс] / Стеганография. Особенности использования программ на основе метода наименьшего значащего бита, 2005. – Режим доступа: <http://www.compdoc.ru/secur/crypt/steganografiya/>, свободный. – Загл. с экрана. – Яз. рус.

КОМПЛЕКСНЫЕ РЕШЕНИЯ В ОБЛАСТИ СОБЛЮДЕНИЯ МЕР ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В БАНКОВСКОЙ СФЕРЕ

М.А. Семёнова, В.А. Семёнов

Научный руководитель – д.т.н., профессор А.Г. Коробейников

В статье рассматриваются политика информационной безопасности организации банковской системы РФ, необходимость учитывать цели организации и имеющиеся ресурсы, необходимость соблюдения стандарта «Обеспечение информационной безопасности организаций банковской системы РФ». Приведены рекомендации соблюдения информационной безопасности в коммерческих банках.

Введение

Под информационной безопасностью (ИБ) банковских систем понимается система мер защиты банка от утечки конфиденциальной информации, в первую очередь о его клиентах. Речь идет о проверке выполнения стандарта «Обеспечение информационной безопасности организаций банковской системы РФ». В документе указывается, что наибольшими возможностями для нанесения ущерба банку обладает его собственный персонал. Сейчас этот стандарт носит рекомендательный характер и необязателен для исполнения. Чтобы разобраться в вопросах применения российских стандартов, необходимо представить себе структуру ИБ и понять, какое место они в ней занимают. Эта структура показана на рис. 1, откуда видно, что стандарты относятся к специальным нормативным документам по технической защите информации и находятся в определенном логическом соответствии с правовыми и организационно-распорядительными документами.



Рис. 1. Структура информационной безопасности в России

Для проведения проверок ЦБ разработал проект стандарта по аудиту информационной безопасности в банках и проект методики оценки ее соответствия требованиям стандарта. Эти документы были представлены для обсуждения в Ассоциацию пользователей стандартов ЦБ по обеспечению ИБ организаций банковской системы России. После того как документы будут утверждены, ими будут руководствоваться аудиторы при проверке ИБ в банках и сами сотрудники ЦБ.

Таким образом, Банк России среагировал на утечки банковской конфиденциальной информации, участвовавшие в последние полгода. С лета 2006 г. в продаже на черном рынке стали появляться базы заемщиков – физических лиц. Источником утечки эксперты и сами банкиры называли службы безопасности банков. Участники рынка опасаются, что проверки выявят огромные объемы утечки конфиденциальной банковской информации, в первую очередь о корпоративных клиентах.

Политика информационной безопасности

Собственник организации должен обеспечить разработку, принятие и внедрение политики ИБ организации БС РФ, включая выделение требуемых для реализации этой политики ресурсов [3].

Для эффективного выполнения целей организации и задач по управлению активами должны быть выделены и определены соответствующие роли персонала организации. Роли следует персонифицировать с установлением ответственности за их исполнение. Формирование ролей, как правило, должно осуществляться на основании бизнес-процессов. Ответственность должна быть зафиксирована в должностных инструкциях [2].

При определении ролей для сотрудников организации БС РФ необходимо учитывать цели организации, имеющиеся ресурсы, функциональные и процедурные требования, критерии оценки эффективности выполнения правил для данной роли. Не рекомендуется, чтобы одна персональная роль целиком отражала цель, например, включала все правила, требуемые для реализации бизнес-процесса. Совокупность правил, составляющих роли, не должна быть критичной для организации с точки зрения последствий успешного нападения на ее исполнителя. Не следует совмещать в одном лице (в любой комбинации) роли разработки, сопровождения, исполнения, администрирования или контроля, например, исполнителя и администратора, администратора и контролера или других комбинаций.

При приеме на работу должны быть проверены идентичность личности, заявляемая квалификация, точность и полнота биографических фактов, наличие рекомендаций.

Лиц, которых предполагается принять на работу, связанную с защищаемыми активами или операциями, следует подвергать проверке в части профессиональных навыков и оценки профессиональной пригодности. Рекомендуется выполнять контрольные проверки уже работающих сотрудников регулярно, а также внепланово при выявлении фактов их нештатного поведения, или участия в инцидентах ИБ, или подозрений в таком поведении или участии.

Весь персонал организации БС РФ должен давать письменное обязательство о соблюдении конфиденциальности, приверженности правилам корпоративной этики, включая требования по недопущению конфликта интересов [3]. При этом условие о соблюдении конфиденциальности должно распространяться на всю защищаемую информацию, доверенную сотруднику или ставшую ему известной в процессе выполнения им своих служебных обязанностей. Для внешних организаций требования по ИБ регламентируются положениями, включаемыми в договоры (соглашения).

Персонал организации должен быть компетентным для выполнения своих функций в области обеспечения ИБ. Компетентность персонала следует обеспечивать с помощью процессов обучения в области ИБ, осведомленности персонала и периодической проверки уровня компетентности.

Привлекаемые для разработки и (или) производства средств и систем защиты АБС на договорной основе специализированные организации должны иметь лицензии на данный вид деятельности в соответствии с законодательством РФ.

Оперативный контроль доступа пользователей осуществляется подразделениями (лицами) в организации, ответственными за обеспечение ИБ; регистрация действий

персонала и пользователей в специальном электронном журнале. Данный электронный журнал должен быть доступным для чтения, просмотра, анализа, хранения и резервного копирования только администратору ИБ. При невозможности поддержки данного режима эксплуатирующимися в организации БС РФ аппаратно-программными средствами реализация данного требования должна быть обеспечена организационными и/или административными мерами.

При обеспечении антивирусной защиты в организации должны быть разработаны и введены в действие инструкции по антивирусной защите, учитывающие особенности банковских технологических процессов. Особое внимание должно быть уделено антивирусной фильтрации трафика электронного почтового обмена.

Лучшей практикой является построение эшелонированной централизованной системы антивирусной защиты, предусматривающей использование средств антивирусной защиты различных производителей и их раздельную установку на рабочих станциях, почтовых серверах и межсетевых экранах.

Электронная почта должна архивироваться. Архив должен быть доступен только подразделению (лицу) в организации, ответственному за обеспечение ИБ. Изменения в архиве не допускаются. Доступ к информации архива должен быть ограничен.

Общие требования по обеспечению информационной безопасности при использовании средств криптографической защиты информации

Средства криптографической защиты информации (СКЗИ) должны допускать встраивание в технологическую схему обработки электронных сообщений, обеспечивать взаимодействие с прикладным программным обеспечением на уровне обработки запросов на криптографические преобразования и выдачи результатов [2].

Внутренний порядок применения СКЗИ в АБС определяется руководством организации и должен включать:

- 1 порядок ввода в действие;
- 2 порядок эксплуатации;
- 3 порядок восстановления работоспособности в аварийных случаях;
- 4 порядок внесения изменений;
- 5 порядок снятия с эксплуатации;
- 6 порядок управления ключевой системой;
- 7 порядок обращения с носителями ключевой информации.

Ключи кодов аутентификации (КА) и/или электронной цифровой подписи (ЭЦП) должны изготавливаться в каждой организации самостоятельно. В случае изготовления ключей КА, ЭЦП для одной организации в другой организации БС РФ согласие первой организации считать данный ключ своим должно быть зафиксировано в договоре. В качестве объектов защиты должны рассматриваться:

- 1 банковский платежный технологический процесс;
- 2 платежная информация;
- 3 технологический процесс по управлению ролями и полномочиями сотрудников организации БС РФ, задействованных в обеспечении банковского платежного технологического процесса.

Банковский платежный технологический процесс должен быть однозначно определен (отражен) в нормативно-методических документах организации БС РФ.

Порядок обмена платежной информацией должен быть зафиксирован в договорах между участниками, осуществляющими обмен платежной информацией. В роли участников могут выступать организации БС РФ, юридические и физические лица.

Сотрудники организации БС РФ, в том числе администраторы автоматизированных систем и средств защиты информации, не должны обладать всей полнотой полномочий

для бесконтрольного создания, авторизации, уничтожения и изменения платежной информации, а также проведения операций по изменению состояния банковских счетов. Результаты технологических операций по обработке платежной информации должны быть контролируемыми (проверены) и удостоверены лицами/автоматизированными процессами. Лица/автоматизированные процессы, осуществляющие обработку платежной информации и контроль (проверку) результатов обработки, должны быть независимы друг от друга.

При работе с платежной информацией необходимо проводить авторизацию и контроль целостности данной информации.

Лучшей практикой при автоматизированной обработке платежной информации является оснащение средств вычислительной техники (на которых осуществляются операции над платежной информацией) сертифицированными или разрешенными руководителем организации БС РФ к применению средствами защиты от НСД и средствами криптографической защиты информации.

Безопасность информации, отнесенной к банковской тайне, обеспечивается в соответствии со статьей 26 Федерального закона «О банках и банковской деятельности» [6].

Администратор ИБ не должен иметь права добавить нового пользователя в АБС, а также удалить из нее существующего пользователя. В случае отсутствия у администратора ИБ технических возможностей по настройке параметров АБС, влияющих на полномочия пользователей по доступу к информации эти настройки выполняются администратором АБС, но с обязательным предварительным согласованием устанавливаемых прав доступа пользователей к информации с администратором ИБ

Для каждой АБС должен быть определен порядок контроля ее функционирования со стороны лиц, отвечающих за ИБ.

Процессы подготовки, ввода, обработки и хранения информации, а также порядок установки, настройки, эксплуатации и восстановления необходимых технических и программных средств должны быть регламентированы и обеспечены инструктивными и методическими материалами, согласованными со службой ИБ. Должна осуществляться и быть регламентирована процедура периодического тестирования всех реализованных программно-техническими средствами функций (требований) по обеспечению ИБ. Регламентирующие документы должны быть согласованы со службой ИБ. Должна осуществляться и быть регламентирована процедура восстановления всех реализованных программно-техническими средствами функций по обеспечению ИБ. Регламентирующие документы должны быть согласованы со службой ИБ.

Рекомендации по проверке соблюдения безопасности информационных систем в коммерческих банках

Банк России ввел новые стандарты обеспечения ИБ банковской системы

Согласно сообщению ЦБ, с целью повышения уровня ИБ организаций банковской системы ЦБ разработал стандарт СТО БР ИББС-1.0. Введение этого стандарта – реакция Банка России на скандалы из-за утечки информации из кредитных организаций. По замыслу ЦБ, ввод в действие этих документов позволит обеспечить единый подход к оценке ИБ российских банков. Кроме того, регулятор считает, что это расширит возможности применения стандарта СТО БР ИББС-1.0 для целей надзора и инспектирования кредитных организаций.

1. Основной целью проведения проверки соблюдения безопасности информационных систем в банках является оценка уровня ИБ банков, оценка рисков нарушений ИБ путем выявления угроз информационной безопасности, определения их источников и возможных последствий для банков. Наиболее опасными (значимыми) угрозами ИБ

информационных систем (способами нанесения ущерба субъектам информационных отношений) являются [3]:

- 1 нарушение конфиденциальности (разглашение, утечка) сведений, составляющих банковскую или коммерческую тайну, а также персональных данных;
- 2 нарушение доступности (дезорганизация работы) информационных систем, блокирование доступа к информации, нарушение технологических процессов, срыв своевременного решения задач;
- 3 нарушение целостности (искажение, подмена, уничтожение) информационных, программных и других ресурсов, а также фальсификация (подделка) документов.

2. Основными объектами проверки являются:

- 1 нормативная база банка по обеспечению ИБ;
- 2 системы обеспечения безопасности информационных ресурсов с ограниченным доступом, составляющих государственную, коммерческую, банковскую тайну, иных чувствительных по отношению к случайным и несанкционированным воздействиям и нарушению их безопасности информационных ресурсов, в том числе открытой (общедоступной) информации, представленной в виде документов и массивов информации, независимо от формы и вида их представления;
- 3 системы обеспечения безопасности процессов обработки информации – информационных технологий, регламента и процедур сбора, обработки, хранения и передачи информации;
- 4 системы обеспечения безопасности информационной инфраструктуры, включающей системы обработки и анализа информации, технических и программных средств ее обработки, передачи и отображения, в том числе каналов информационного обмена и телекоммуникаций, объектов и помещений, в которых размещены компоненты инфраструктуры.

3. Основными источниками угроз ИБ являются [6]:

- 1 непреднамеренные (ошибочные, случайные, необдуманные, без злого умысла и корыстных целей) нарушения установленных регламентов сбора, обработки и передачи информации, а также другие действия персонала при эксплуатации информационных систем, приводящие к непроизводительным затратам времени и ресурсов, разглашению сведений ограниченного распространения, потере ценной информации или нарушению работоспособности отдельных рабочих станций, подсистем или в целом всей системы;
- 2 преднамеренные (в корыстных целях, по принуждению третьими лицами, со злым умыслом и т.п.) действия сотрудников, допущенных к работе с информационными системами, а также сотрудников, отвечающих за обслуживание, администрирование программного и аппаратного обеспечения, средств защиты и обеспечения информационной безопасности;
- 3 деятельность преступных групп и формирований, политических и экономических структур, а также отдельных лиц по добыванию информации, навязыванию ложной информации, нарушению работоспособности системы в целом и ее отдельных компонентов;
- 4 ошибки, допущенные при проектировании информационных систем и их систем защиты, ошибки в программном обеспечении, отказы и сбои технических средств (в том числе средств защиты информации и контроля эффективности защиты);
- 5 действия компьютерных вирусов;
- 6 аварии, стихийные бедствия и т.п.

4. Нормативная база по обеспечению ИБ банка должна, как минимум, включать список нормативных документов, приведенных в приложении к Стандартам по обеспечению информационной безопасности учреждений банковской системы и соответствовать нормативным актам. Также инспектору необходимо изучить контрольную отчет-

ность – регистрационные журналы и отчеты по безопасности информационных систем, все организационно-распорядительные документы по вопросам обеспечения ИБ, а также отчеты внутреннего и внешнего аудита информационных систем.

5. В процессе проведения проверки инспектору необходимо изучить структуру информационной системы банка, а также нормативные документы, регламентирующие работу с информационными системами банка, а именно:

- 1 организация локальной вычислительной сети (ЛВС) в банке;
- 2 способы подключения к ЛВС рабочих мест сотрудников банка;
- 3 организация резервирования и восстановления важной информации. Организация восстановления работоспособности информационных систем на случай возникновения чрезвычайных ситуаций;
- 4 организация подключения к ресурсам сети Интернет и электронной почте;
- 5 способы защиты электронных платежей;
- 6 организация безопасности внутреннего электронного документооборота в банке;
- 7 организация делопроизводства с конфиденциальными документами. Доступ лиц к работе с конфиденциальными документами;
- 8 организация взаимодействия банка с представительствами банка и филиалами;
- 9 организация обучения сотрудников правилам безопасности;
- 10 организация систем архивирования, систем обнаружения атак злоумышленников и антивирусной защиты.

6. Проверка обеспечения безопасности при использовании пластиковых карт
В целях проверки обеспечения безопасности при использовании пластиковых карт инспектору необходимо проверить:

- 1 обеспечение контроля над целостностью и подлинностью передачи транзакций по каналам связи, а также последующей идентификацией и аутентификацией держателя карты;
- 2 проведение мониторинга и контроля над транзакциями в системе, мониторинга и управления сетью терминалов и банкоматов;
- 3 применение методов шифрования и верификации.

7. Проверка обеспечения безопасности коммуникаций. В целях проверки обеспечения безопасности коммуникаций инспектору необходимо проверить:

- 1 обеспечение безопасности использования линий связи и коммуникационного оборудования;
- 2 наличие резервных линий связи или альтернативных способов передачи информации;
- 3 применение средств криптографии, межсетевое экранирование данных, а также средств аутентификации при защите информации на внешнем уровне;
- 4 обеспечение защиты платежных документов;
- 5 обеспечение регистрации и хранения платежных документов.

8. Проверка обеспечения кадровой безопасности. В целях проверки обеспечения кадровой безопасности инспектору необходимо проверить:

- 1 наличие соответствующего персонала;
- 2 применение нормативной политики;
- 3 соблюдение сегрегации полномочий;
- 4 обеспечение контроля над установлением и соблюдением полномочий пользователей.

9. Проверка обеспечения целостности данных. В целях проверки обеспечения целостности данных инспектору необходимо проверить:

- 1 обеспечение дублирования данных;
- 2 обеспечение периодического резервного копирования и восстановления данных;
- 3 обеспечение соответствующей защитой данных на внешнем уровне.

10. Проверка обеспечения конфиденциальности данных. В целях проверки обеспечения конфиденциальности данных инспектору необходимо проверить:

- 1 применение нормативной политики безопасности;
- 2 обеспечение регистрирования доступа к конфиденциальным данным;
- 3 применение специальных средств кодирования информации (средства криптографии) и межсетевое экранирование.

11. Проверка контроля доступа к информационным ресурсам. В целях проверки недопущения несанкционированного доступа к информационным ресурсам инспектору необходимо проверить:

- 1 наличие системы разграничения доступа сотрудников к работе с автоматизированными системами;
- 2 применение регистрации пользователя, его действий, а также несанкционированных операций;
- 3 применение идентификации пользователя;
- 4 применение антивирусной защиты.

12. Проверка обеспечения безопасности программного обеспечения. В целях выявления нарушений соблюдения безопасности автоматизированных систем (АС), инспектору необходимо проверить:

- 1 соблюдение ограничения доступа к программному обеспечению;
- 2 разграничение прав и полномочий при работе с АС;
- 3 наличие резервной копии программного обеспечения.

13. Проверка обеспечения безопасности технических средств. В целях выявления нарушений соблюдения безопасности технических средств инспектору необходимо проверить [1]:

- 1 наличие помещений для оборудования информационной системы; соответствующего условиям эксплуатации данного оборудования;
- 2 соответствие технической укрепленности помещений требованиям НБКР;
- 3 соблюдение ограничения доступа в специальные помещения;
- 4 наличие специального оборудования для хранения носителей информации, обеспечивающего защиту от несанкционированного доступа, от теплового, механического и электромагнитного воздействия;
- 5 использование автономных и бесперебойных источников питания.

14. Составление отчета о проверке. По окончании проверки инспектор составляет отчет о проверке обеспечения безопасности информационных систем, включающий, как минимум, следующее [5]:

- 1 схематичное описание структуры информационной системы банка;
- 2 описание результатов проверки всех вышеизложенных разделов данных методических рекомендаций;
- 3 рекомендации по устранению нарушений соблюдения безопасности информационных систем, если таковые были выявлены.

Заключение

Информационная безопасность банковских систем – это система мер защиты банка от утечки конфиденциальной информации, выполнение стандарта «Обеспечение информационной безопасности организаций банковской системы РФ». Проанализированы необходимые меры по проверке соблюдения безопасности информационных систем в коммерческих банках и определены пути их устранения.

Литература

1. Галатенко В.А. Основы информационной безопасности. – М.: ИНТУИТ.РУ, 2003. – 280 с.

2. Закон Российской Федерации «Об информации, информатизации и защите информации» от 25.01.95 // Собрание законодательства Российской Федерации. – 1995. – № 8. – С. 609.
3. Стрельченко Ю.А. Обеспечение информационной безопасности банков: Метод. пособие – М.: ИПКИР, 1994.
4. Шаваев А.Г. и др. Экономическая безопасность. Энциклопедия. – М.: Издательский дом «Правовое просвещение, 2001.
5. Шпак В. Методологические основы обеспечения информационной безопасности объекта // Конфидент. Защита информации. – 2006, январь-февраль, – С. 75–86.
6. Стандарт «Обеспечение информационной безопасности организаций банковской системы РФ. Общие положения» // Вестник Банка России. – № 68 (792). – 24 ноября 2004.

ОЦЕНКА УСТОЙЧИВОСТИ ЦВЗ К ВНЕШНИМ ВОЗДЕЙСТВИЯМ, ВНЕДРЕННЫХ С ПОМОЩЬЮ АЛГОРИТМОВ ПРОСТРАНСТВЕННОЙ ОБЛАСТИ ВСТРАИВАНИЯ

О.В. Михайличенко, Н.Н. Прохожев

Научный руководитель – д.т.н., профессор А.Г. Коробейников

Целью работы является изучение устойчивости внедренной информации при различных внешних воздействиях на изображение-контейнер. В рамках исследований контейнеры с внедренной информацией подвергаются ряду наиболее распространенных манипуляций с изображениями. Таким образом, в работе получены показатели устойчивости алгоритмов пространственной области к различного рода атакам, которые помогут в выборе алгоритмов при проектировании стеганосистемы.

Введение

Стеганографические алгоритмы пространственной области являются одними из самых ранних алгоритмов, использовавшихся для сокрытия информации в неподвижных изображениях. Появившись на заре развития современной цифровой стеганографии, они обладают неоспоримым достоинством, а именно – математической простотой реализации, не требующей больших вычислительных мощностей. Однако к настоящему времени данное преимущество уже не столь актуально в силу многократно возросших вычислительных мощностей современной компьютерной техники. Одним из важнейших требований, предъявляемых к стеганографическим алгоритмам, является их устойчивость к внешним воздействиям. Данное требование особенно актуально для алгоритмов внедрения цифровых водяных знаков, или ЦВЗ [1].

В работе проведен анализ одного из наиболее удачных, с точки зрения авторов статьи, стеганографических алгоритмов пространственной области – алгоритма Лангелаар. Цель работы – оценка устойчивости алгоритмов пространственной области встраивания, на примере алгоритма Лангелаар, к внешним воздействиям или атакам на изображение-контейнер.

Алгоритм встраивания и извлечения ЦВЗ

Являясь типичным представителем своего класса, алгоритм Лангелаар [2] осуществляет внедрение скрываемой информации непосредственно в пиксели изображения. Механизм внедрения можно описать следующими шагами.

- (1) Изображение разбивается на блоки 8×8 пикселей.
- (2) С помощью маски, сгенерированной случайным образом, пиксели блока делятся на две группы, равные по количеству пикселей.
- (3) Вычисляется среднее значение яркости пикселей в каждой группе l_0 и l_1 .
- (4) Встраиваемый бит однозначно кодируется по формулам (1) и (2):

$$l_0 - l_1 > +P, \text{ если } s = 1; \quad (1)$$

$$l_0 - l_1 > -P, \text{ если } s = 0, \quad (2)$$

где P – коэффициент силы встраивания, s – встраиваемый бит. В качестве ЦВЗ для наглядности использовано битовое изображение, хотя более традиционной для данного алгоритма форма ЦВЗ – это битовая строка символов.

Оценка устойчивости встроенной информации к внешним воздействиям

Определим, как внедрение в различные частотные области коэффициентов отражается на таком важном свойстве, как устойчивость внедренной информации к внешним воздействиям. Технология определения устойчивости того или иного алгоритма или стеганосистемы в целом состоит из следующих шагов.

- (1) Скрываемая информация внедряется в контейнер.
- (2) Контейнер подвергается внешнему воздействию или атаке.

(3) Скрытая информация извлекается из контейнера.

(4) Извлеченная информация сравнивается с оригинальной, и определяется степень их соответствия.

Оценку устойчивости предлагается выполнять с помощью простого коэффициента ошибочных бит (Bit Error Rate). Коэффициент ошибочных бит прекрасно подходит для оценки искажений в битовой последовательности, которую в данном случае и представляет собой скрываемая информация. Вычисляется данный коэффициент по формуле (3):

$$\text{BER}(S, S'') = \frac{\sum p_i}{N}, \quad (3)$$

где $p_i = \begin{cases} 1 & \text{если } s_i \neq s_i'' \\ 0 & \text{если } s_i = s_i'' \end{cases}$,

s_i – i -й бит оригинала встраиваемой строки; s_i'' – бит извлеченной строки; N – общее количество бит. Таким образом, при значении коэффициента, равном 0, внедренная и извлеченная информация полностью идентичны. При значении, равном 1, каждый бит оригинала не соответствует извлеченному, т.е. мы имеем дело с «негативом». Исходя из этого, при значении, близком к 0,5, можно с уверенностью утверждать, что две битовые строки совершенно некоррелированы друг с другом.

На рис. 1 представлены ЦВЗ, поврежденные в результате внешнего воздействия на изображение-контейнер, с различными значениями BER. Можно с полной уверенностью констатировать, что при значении BER более 0.3 можно говорить о полной непригодности встроенной графической информации. Совершенно очевидно то, что для ЦВЗ, представленным битовой строкой символов, это значение гораздо меньше. При 8-и битовом кодировании символа даже несоответствие всего одного бита в каждом символе, что при значении BER, равном 0.125, приведет к полной потере информации.

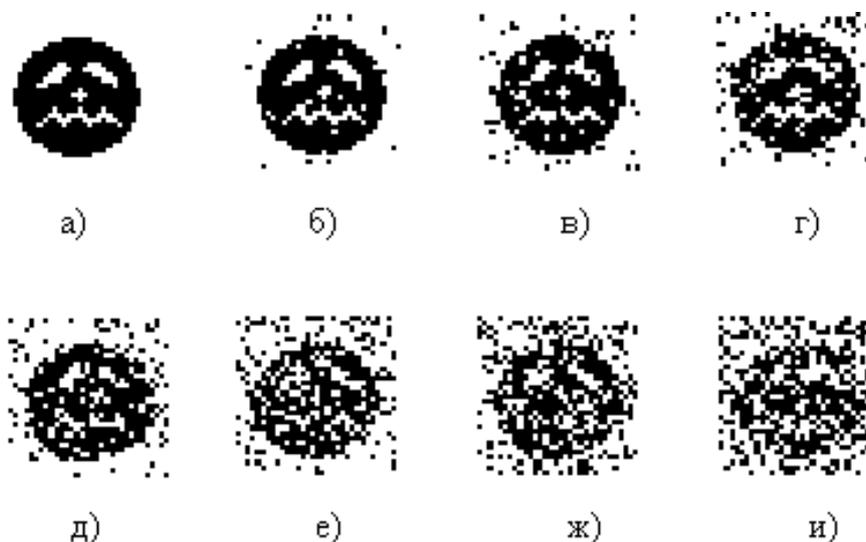


Рис. 1. ЦВЗ, поврежденные в результате внешних атак на контейнер:
а) оригинал; б) BER= 0.03; в) BER= 0.07; г) BER= 0.1; д) BER= 0.15;
е) BER= 0.20, ж) BER= 0.25, и) BER= 0.30

Внешние воздействия на изображение-контейнер

В качестве внешних воздействий применим сжатие JPEG, зашумление изображения-контейнера белым гауссовским шумом, масштабирование, фильтрация. Все эти воз-

действия оказывают довольно сильные и, что более важно, необратимые воздействия на контейнер, приводящие к значительным искажениям, и, следовательно, представляют реальную угрозу для встроенной в контейнер информации. Исследования проводились на 10 естественных полутоновых изображениях размером 512×512 пикселей. В силу того, что устойчивость встроенной информации напрямую зависит от величины коэффициента силы встраивания, все исследования проводились с фиксированным значением этого параметра. Значение коэффициента выбиралось исходя из условия отсутствия каких-либо визуально определяемых артефактов внедрения. Экспериментальным путем было выбрано значение $P = 10$.

Сжатие JPEG с потерями

Для проверки устойчивости к сжатию JPEG изображение-контейнер подвергалось сжатию JPEG во всем диапазоне значений коэффициента качества JPEG. Результат представлен на рис. 2. Как можно видеть из графика, алгоритм явно не обладает достаточной устойчивостью к такого рода внешним воздействиям, поскольку естественные изображения, сжатые с коэффициентом качества 70–60, визуально несильно деградированы и могут быть использованы в коммерческих целях.

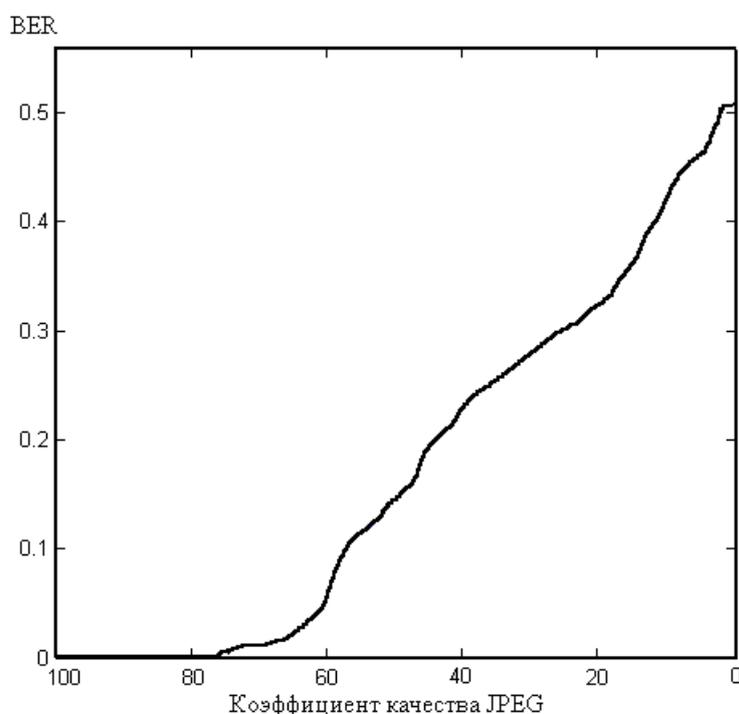


Рис. 2. Устойчивость встроенной информации к сжатию JPEG с потерями

Гауссовский белый шум

Для проверки устойчивости к зашумлению в контейнер вносился белый гауссовский шум с нулевым средним значением и разными значениями отклонения, изменяющимися от «0» в сторону возрастания до тех пор, пока деградация изображения не достигла уровня, совершенно неприемлимого для использования в коммерческих целях. Результат представлен на рис. 3.

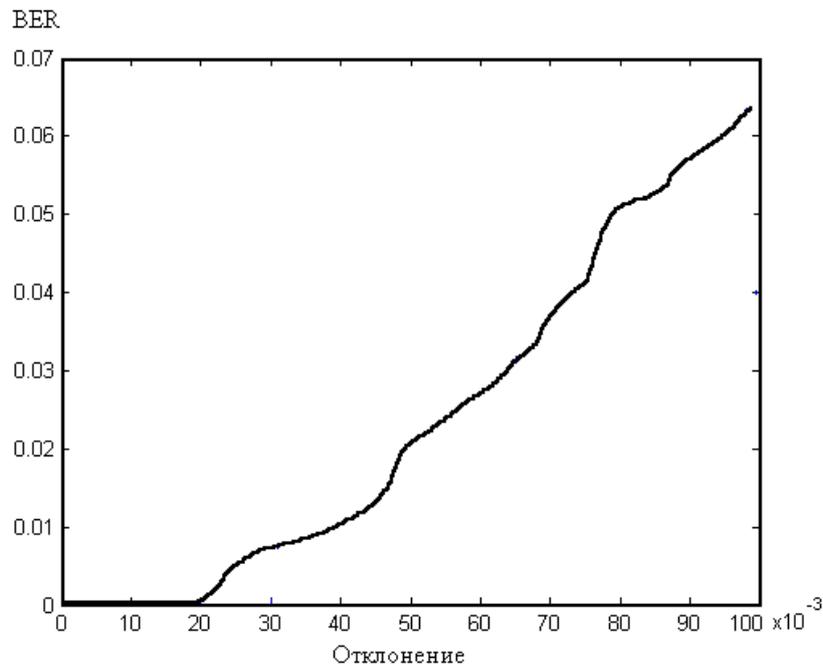


Рис. 3. Устойчивость встроенной информации к зашумлению

Масштабирование

В ходе эксперимента с масштабированием изображение-контейнер сжималось до различных размеров вплоть до 90% сжатия, т.е. в 10 раз. Совершенно очевидно, что считывать ЦВЗ из сжатого изображения неприемлемо, по крайней мере, не уменьшив пропорционально размер блоков, на которые разбивается изображение. Однако такой задачи и не ставилось, поскольку изображение всегда можно восстановить до размеров исходного, а уж затем попытаться считать ЦВЗ. Это было проделано в ходе эксперимента, результаты которого представлены на рис. 4.

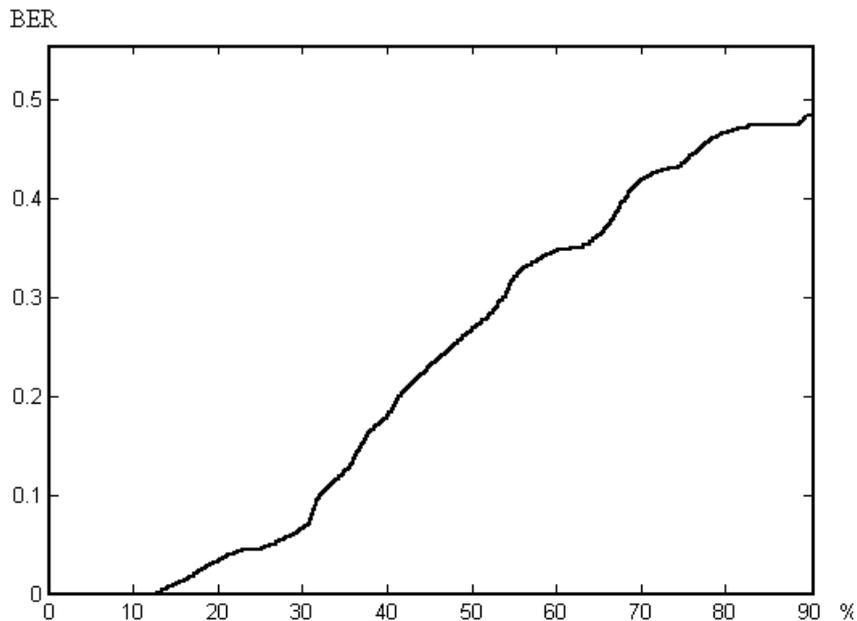


Рис. 4. Устойчивость встроенной информации к масштабированию

Применение фильтров

Фильтрацию также можно отнести к разряду наиболее вероятных внешних воздействий на контейнер с внедренным ЦВЗ. Из большого многообразия фильтров для

цифровых изображений были выбраны три вида фильтров: низкочастотный гауссовский фильтр, усредняющий фильтр с размером окна 3×3 пикселя и контрастный фильтр, повышающий резкость изображения с тем же размером окна. Результаты эксперимента представлены в таблице.

Фильтр	Значение BER
Низкочастотный	0.16
Усредняющий	0.44
Контрастный	0

Таблица. Устойчивость встроенной информации к фильтрации

Заключение

Алгоритм демонстрирует отличные показатели устойчивости к зашумлению изображения-контейнера и неплохую устойчивость к масштабированию. Однако к внешним воздействиям частотной области, таким как усредняющая фильтрация и сжатие JPEG с потерями, данный алгоритм совершенно неустойчив, что, разумеется, сильно ограничивает область его применения. Надо сказать, что автор этого алгоритма понимал необходимость повышения устойчивости к сжатию JPEG и разработал его модификацию, использующую предварительное дискретное косинусное преобразование изображения, что, разумеется, повышает устойчивость, но значительно усложняет реализацию и повышает его вычислительную емкость.

Таким образом, можно рекомендовать данный вид алгоритмов исключительно для изображений-контейнеров, для которых вероятность атаки в виде сжатия JPEG с низкими коэффициентами качества крайне маловероятна.

Литература

1. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография: Теория и практика. – М.: МК-Пресс, 2006. – 283 с.
2. Gerrit C. Langelaar, Reginald L. Lagendijk, and Jan Biemond. Robust labeling methods for copy protection of images // Proceedings of the SPIE Storage and Retrieval for Image and Video Databases V. – Volume 3022. – San Jose, California, 1997.

ИНФОРМАЦИОННАЯ ВОЙНА. ИНФОРМАЦИОННОЕ ОРУЖИЕ. СОВРЕМЕННАЯ СИТУАЦИЯ

А.А. Кривошеев

Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В статье рассматриваются основные понятия информационной войны, информационного оружия; современная мировая ситуация в информационной среде.

Стремительные темпы развития компьютеризации и информатизации общества неизбежно ведут к созданию единого мирового информационного пространства, в котором будут аккумулированы все средства сбора, накопления, обработки, обмена и хранения информацией. В России современная информационная инфраструктура в настоящее время только формируется, и входящие в нее информационные системы зачастую не имеют выхода в открытые сети связи. В связи с этим опасности для информационной инфраструктуры России кажутся преувеличенными. Однако по мере внедрения информационных технологий в различные сферы общественной жизни страны возникают серьезные проблемы, связанные с обеспечением исправного функционирования элементов информационной инфраструктуры, сохранности информации и т.д., с которыми уже столкнулись в развитых странах Запада и, в первую очередь в США, обладающих половиной всего мирового информационного ресурса.

В России уже появился ряд концептуальных документов, рассматривающих информационные угрозы национальной безопасности. Прежде всего, это Концепция национальной безопасности Российской Федерации и Доктрина информационной безопасности Российской Федерации. Так, в Концепции четко определены угрозы стране в информационной сфере [1]. Это «стремление ряда стран к доминированию в информационном пространстве, вытеснению России с внешнего и внутреннего информационного рынка; разработкой рядом государств концепции информационных войн, предусматривающей создание средств опасного воздействия на информационные сферы других стран мира; нарушение нормального функционирования информационных и телекоммуникационных систем, а также сохранности информационных ресурсов, получение несанкционированного доступа к ним» [2].

Информационное пространство фактически стало театром военных действий, где каждая противоборствующая сторона стремится получить преимущество, а в случае необходимости разгромить противника. Размах противоборства в информационной сфере достиг таких масштабов, что потребовалось создание специальной концепции, получившей название «информационной войны» или «информационного противоборства». Впервые работы по созданию концепции «информационной войны» начались в США в начале 90-х гг. В настоящее время существует несколько вариантов трактовки термина «информационная война». Однако отличия между ними незначительны, поэтому с полным основанием можно использовать вариант термина, представленного в Уставе Сухопутных войск США FM 100-6 «Информационные операции» (август 1996 г.). Согласно этому документу, «Информационная война – это комплекс мероприятий по достижению информационного превосходства путем воздействия на информацию, информационные процессы, информационные системы и компьютерные сети противника при одновременной защите своей информации, информационных процессов, информационных систем и компьютерных сетей».

В рамках информационной войны проводятся мероприятия наступательного и оборонительного характера. Соответственно, совершенствуются уже существующие и активно разрабатываются новые оборонительные и наступательные средства ведения информационного противоборства, которые позволят достичь информационное превосходство над противником.

Чтобы определить, имеет ли понятие «информационное оружие» право на существование, следует, прежде всего, обратиться к определению оружия [1]. В Советской военной энциклопедии оружие определяется как «устройства и средства, применяемые в вооруженной борьбе для поражения и уничтожения противника» [3]. Основным в определении оружия представляется цель его использования, а именно поражение противника. Под поражением объектов (целей) подразумевается воздействие различными средствами поражения на объекты (цели), в результате которого они полностью или частично (временно) теряют способность к нормальному функционированию (выполнению боевой задачи).

Поражение объектов заключается в их *уничтожении (разрушении), подавлении и изнурении* (живой силы объектов). Уничтожение объекта состоит в причинении ему такого ущерба, при котором он полностью теряет боеспособность. Подавление предполагает нанесение объекту такого ущерба (повреждений) и создание для него таких условий, при которых он временно лишается боеспособности, ограничивается (воспрещается) его маневр или нарушается управление. Изнурение состоит в длительном ведении по объекту беспокоящего огня ограниченным количеством сил и средств или нанесении по нему периодических ударов авиации. Основной его целью является морально-психологическое воздействие на живую силу объекта и тем самым снижение его боеспособности и воспрещение нормального функционирования.

Итак, способно ли так называемое информационное оружие поражать противника? Информационное оружие, согласно одному из существующих определений – это комплекс программных и технических средств, предназначенных для контроля информационных ресурсов объекта воздействия и вмешательства в работу его информационных систем. Информационное оружие можно классифицировать по методам воздействия на информацию, информационные процессы и информационные системы противника. Это воздействие может быть физическим, информационным, программно-техническим или радиоэлектронным.

Физическое воздействие может быть осуществлено путем применения любых средств огневого поражения. Однако более корректным было бы отнести к информационному оружию физического воздействия средства, предназначенные исключительно для воздействия на элементы информационной системы: противорадиолокационные ракеты, специализированные аккумуляторные батареи генерации импульса высокого напряжения, средства генерации электромагнитного импульса, графитовые бомбы, биологические и химические средства воздействия на элементарную базу.

С помощью противорадиолокационных ракет в первые дни воздушной операции коалиционных миротворческих сил в зоне Персидского залива (1991 г.) было выведено из строя 80% наземных РЛС Ирака. Эффективным является и использование генераторов электромагнитного излучения. Проведенные эксперименты показали, что простейшие малогабаритные генераторы на расстоянии до 500 м могут внести опасные искажения в работу приборов самолета, совершающего взлет или посадку, а также заглушать двигатели современных автомобилей, оснащенных микропроцессорной техникой. Графитовые бомбы применялись американскими вооруженными силами в ходе войны в Персидском заливе и в Косово. Их поражающий эффект достигается путем создания над объектом облака площадью до 200 м² из произведенных на основе углерода и обладающих сверхпроводимостью тонких волокон. При соприкосновении волокон с токонесущими элементами (изоляторы, провода и т.д.) происходило короткое замыкание и вывод из строя электроцепей.

Биологические средства представляют собой особые виды микробов, способные уничтожать электронные схемы и изолирующие материалы, используемые в радиоэлектронной технике.

Информационные методы воздействия реализуются посредством всей совокупности средств массовой информации и глобальных информационных сетей типа Интернет, станциями голосовой дезинформации.

Так как основным элементом информационной инфраструктуры являются люди, мотивация деятельности которых базируется на их физиологических, социальных и информационных потребностях, то правильно рассчитанное применение так называемых информационно-психологических методов воздействия оказывает прямое влияние на уровень безопасности государства. Особенно это касается России, где пока отсутствует организованная система формирования и поддержания в обществе необходимых нравственных ценностей, патриотизма и гражданской ответственности за судьбу страны. Научно-технический прогресс в области информационных технологий, развитие СМИ стерли национальные границы в информационном пространстве и создали беспрецедентные возможности для подавления противника с помощью нетрадиционных средств поражения, не вызывающих физического разрушений. Проходя через сознание каждого члена общества, длительное массированное информационно-психологическое воздействие разрушающего характера создает реальную угрозу существованию нации в результате трансформации ее исторически сложившейся культуры, основных мировоззренческих и идеологических установок.

Существуют многочисленные факты, когда не только зарубежные, но и некоторые российские средства массовой информации, прикрываясь лозунгами «объективности информационного освещения» тех или иных событий, наносят ущерб информационной безопасности страны путем манипулирования информацией, распространения дезинформации, информационной поддержкой различных группировок экстремистского и криминального толка. И эта проблема усугубляется как за счет монополизации отечественных СМИ, так и за счет неконтролируемого расширения сектора зарубежных СМИ в информационном пространстве России.

Станции голосовой дезинформации, разрабатываемые в настоящее время в США, позволят входить в радиосети объекта воздействия и смоделированным компьютерными средствами голосом командира подразделения (части) противника отдавать приказы и распоряжения подчиненным им войскам, тем самым нарушая управление войсками.

Средствами реализации программно-технических методов являются компьютерные вирусы, логические бомбы и аппаратные закладки, а также специальные средства проникновения в информационные сети. Данные средства используются для сбора, изменения и разрушения информации, хранящейся в базах данных, а также для нарушения или замедления выполнения различных функций информационно-вычислительных систем. Программно-технические средства можно классифицировать согласно выполняемым с их помощью задачам на средства сбора информации, средства искажения и уничтожения информации и средства воздействия на функционирование информационных систем. Причем некоторые средства могут быть универсальными и использоваться как для искажения (уничтожения) информации, так и для воздействия на функционирование информационных систем объекта воздействия.

Средства сбора информации позволяют производить несанкционированный доступ к компьютерным системам, определять коды доступа, ключи к шифрам или другую информацию о зашифрованных данных и по каналам обмена передавать полученные сведения заинтересованным организациям. В настоящее время разработаны специальные программные продукты, так называемые «ноуботы» (Knowbot – Knowledge Robot), которые способны перемещаться в информационной сети от компьютера к компьютеру и при этом размножаться, создавая копии. «Ноубот» вводится в компьютерные системы и, обнаружив интересующую его информацию, оставляет в этом месте свою копию, которая собирает информацию и в определенное время передает ее. С целью исключения

обнаружения в «ноуботе» могут быть предусмотрены функции самоперемещения и самоуничтожения.

Задачи сбора информации решаются и с помощью программных продуктов «Демон» (Demon), «Вынюхиватели» (Sniffers), «Дверь-ловушка» (Trap Door). Программный продукт «Демон», введенный в систему, записывает все команды, вводимые в нее, и в определенное время передает информацию об этих командах. Аналогично действуют и «Вынюхиватели», которые считывают и передают первые 128 битов информации, необходимых для входа в систему. Программы используются для вскрытия кодов доступа и шифров. «Дверь-ловушка» позволяет осуществлять несанкционированный доступ к информационным массивам базы данных в обход кодов защиты. При этом система и элементы защиты его не распознают.

Созданы и постоянно модернизируются специальные технические устройства, позволяющие считывать информацию с мониторов компьютеров. Перспективным является также создание миниатюрных специализированных комплексов сбора, обработки и передачи информации, которые могут внедряться под видом обычных микросхем в состав самых различных радиоэлектронных устройств.

Средства искажения и уничтожения информации включают программные продукты «Троянский конь» (Trojan Horse), «Червь» (Worm), а также многочисленные компьютерные вирусы, количество которых превышает 60 тысяч. «Троянский конь» позволяет осуществить скрытый несанкционированный доступ к информационным массивам. Он активируется по команде и используется для изменения или разрушения информации, а также замедлять выполнение различных функций системы. «Червь» – это посторонний файл, сформированный внутри информационной базы данных системы. Он способен изменять рабочие файлы, уменьшать ресурсы памяти, а также перемещать и изменять определенную информацию.

К средствам воздействия на функционирование информационных систем относятся «Логические бомбы», «Бомбы электронной почты» и т.д. Логическая бомба представляет собой инструкцию, находящуюся в неактивном состоянии до получения команды на выполнение определенных действий на изменение или разрушение данных, а также нарушение работоспособности информационно-вычислительных систем. Так, в ходе войны в Персидском заливе Ирак не смог применить против многонациональных сил закупленные во Франции системы ПВО, так как их программное обеспечение содержало логические бомбы, активизированные с началом боевых действий.

Бомбы электронной почты – это большой объем несанкционированных сообщений с целью увеличения нагрузки на сервер, чтобы он стал недоступен или его ресурсы стали недостаточными для нормальной работы. Именно таким образом был заблокирован в марте 1999 г. на трое суток сервер НАТО. Неизвестный адресат регулярно присылал на адрес Североатлантического блока около 2000 телеграмм в день, которые переполнили электронный «почтовый ящик».

Радиоэлектронные методы воздействия предполагают использования средств радиоэлектронного подавления, радиоэлектронной разведки и некоторые другие. Основным предназначением такого оружия является контроль информационных ресурсов потенциального противника и скрытое или явное вмешательство в работу его систем управления и связи в целях дезорганизации, нарушения нормального функционирования или вывода их из строя как в мирное, так и в военное время при действиях самостоятельно либо в сочетании с другими средствами воздействия на противника.

Таким образом, практически все вышеперечисленные средства ведения информационного противоборства используются для поражения объектов (целей противника), в основном с целью их подавления. Вопрос вызывает лишь причисление к оружию средств массовой информации, включая глобальную информационную сеть Интернет, а также программно-технические и радиоэлектронные средства сбора информации.

Что касается средств массовой информации, то использование их с целью оказания активного информационно-психологического воздействия может снизить или даже лишить личный состав противника на определенный период боеспособности, заставив его уклоняться различными путями от участия в боевых действиях. В этом случае СМИ выступают в качестве средства подавления, т.е. относятся к оружию.

Программно-технические и радиоэлектронные средства сбора информации не попадают под классическое определение оружия, так как они не участвуют в непосредственном поражении противника, а лишь обеспечивают условия для эффективного ведения вооруженного и, в частности, информационного противоборства. Но если принять за основу сформулированное выше определение информационного оружия, то средства сбора информации, несомненно, обеспечивают контроль над информационными ресурсами противника и могут быть причислены к этому виду оружия.

Основными способами применения информационного оружия могут быть:

- нанесение ущерба отдельным физическим элементам информационной инфраструктуры (разрушение сетей электропитания, создание помех, использование специальных программ, стимулирующих вывод из строя аппаратных средств, а также биологических и химических средств разрушения элементарной базы);
- уничтожение или повреждение информационных, программных и технических ресурсов противника, преодоление систем защиты, внедрение вирусов, программных закладок и логических бомб;
- воздействие на программное обеспечение и базы данных информационных систем и систем управления с целью их искажения или модификации;
- угроза или проведение террористических актов в информационном пространстве (раскрытие и угроза обнародования или обнародование конфиденциальной информации об элементах национальной информационной инфраструктуры, общественно значимых и военных кодах шифрования, принципов работы систем шифрования, успешного опыта ведения информационного терроризма и др.);
- захват каналов СМИ с целью распространения дезинформации, слухов, демонстрации силы и доведения своих требований;
- уничтожение и подавление линий связи, искусственная перегрузка узлов коммутации;
- воздействие на операторов информационных и телекоммуникационных систем с использованием мультимедийных и программных средств для ввода информации в подсознание или ухудшения здоровья человека;
- воздействие на компьютерное оборудование боевой техники и вооружений с целью вывода их из строя [1].

Универсальность, скрытность, многовариантность форм программно-аппаратной реализации, радикальность воздействия, достаточный выбор времени и места применения, наконец, экономичность делают информационное оружие чрезвычайно опасным: оно легко маскируется под средства защиты, например, интеллектуальной собственности; оно позволяет даже вести наступательные действия анонимно, без объявления войны.

Нормальная жизнедеятельность общественного организма целиком определяется уровнем развития, качеством функционирования и безопасностью информационной среды. Производство и управление, оборона и связь, транспорт и энергетика, финансы, наука и образование, средства массовой информации – все зависит от интенсивности информационного обмена, полноты, своевременности, достоверности информации. Именно информационная инфраструктура общества – мишень информационного оружия. Но в первую очередь новое оружие нацелено на вооруженные силы, предприятия оборонного комплекса, структуры, ответственные за внешнюю и внутреннюю безопасность страны. Высокая степень централизации структур государственного управления

российской экономикой может привести к губительным последствиям в результате информационной агрессии. Темпы совершенствования информационного оружия (как, впрочем, и любого вида атакующего вооружения) превышают темпы развития технологий защиты. Поэтому задача нейтрализации информационного оружия, отражения угрозы его применения должна рассматриваться как одна из приоритетных задач в обеспечении национальной безопасности страны [4].

Пару лет назад Центральное разведывательное управление (ЦРУ) упоминало только Россию и Китай в качестве основных источников угрозы из киберпространства. Сегодня американские эксперты отмечают, что уже более 20 стран планируют и осуществляют различные виды информационных операций, направленных против Соединенных Штатов. ЦРУ отмечает, что ряд противостоящих США государств, включают информационную войну как часть их новых военных доктрин. Рассекреченная оценка угрозы, проведенная Военно-морским флотом США, выделяет Россию, Китай, Индию и Кубу в качестве стран, которые открыто, подтвердили политику подготовки к информационной войне и быстро развивают их способности в этом направлении. Северная Корея, Ливия, Иран, Ирак и Сирия, по сообщениям, имеют некоторую способность к движению в этом направлении, а Франция, Япония и Германия уже весьма активны в этой области.

Россия. Информационная инфраструктура России в виду ряда причин является уязвимой от воздействия наступательных средств ведения информационного противоборства, получивших название «информационное оружие» [5]. К этим причинам Доктрина информационной безопасности Российской Федерации относит:

- отсутствие единой государственной политики в области обеспечения информационной безопасности России;
- недостаточное финансирование мероприятий по обеспечению информационной безопасности России;
- увеличение технологического отрыва ведущих держав мира и наращивание их возможностей по противодействию созданию конкурентоспособных российских информационных технологий;
- монополизация информационного рынка России, его отдельных секторов отечественными и зарубежными информационными структурами;
- использование несертифицированных отечественных и зарубежных информационных технологий, средств защиты информации, средств информатизации, телекоммуникации и связи при создании и развитии российской информационной инфраструктуры и т.д.

В отличие от подхода, обозначенного США, в российской Доктрине на первое место ставится обеспечение информационной безопасности индивидуального, группового и общественного сознания. Для реализации основных положений Доктрины и обеспечения информационной безопасности России было создано Управление информационной безопасности в Совете безопасности РФ [6]. Сегодня в работах по разработке отечественного представления информационной войны занимаются Министерство обороны, ФАПСИ, ФСБ и знаменитое Управление «Р» МВД, которое проводит расследования преступлений в высокотехнологической сфере информационных технологий.

США. Деятельность американской администрации в области защиты критической инфраструктуры берет свое начало с формирования Президентской комиссии по защите критической инфраструктуры в 1996 г. Во исполнение указаний президента, обозначенных в этом направлении, был разработан Национальный план защиты информационных систем США, подписанный президентом 7 января 2000 г. На реализацию этого плана было затребовано 2,03 миллиарда долларов из федерального бюджета. С целью улучшения способности активно защищать информационные системы и компьютеры была создана Объединенная оперативная группа по защите компьютерной сети Мини-

стерства обороны, а главнокомандующий космического командования принял полную ответственность за защиту сетей ЭВМ министерства с 1 октября 1999 г. С 2000 г. Министерством обороны начата работа с союзниками по вопросу обеспечения информационной безопасности: Канада имеет официального представителя, работающего в США, развивается система разделения информации между Министерствами обороны в соответствии с основными положениями Меморандума о понимании и Концепции действий подписанными с канадской стороны. Тем не менее, отмечает эксперт, информация, доступная по каналам Интернета, позволяет сделать вывод о том, что уровень информационной безопасности систем Минобороны США, несмотря на реализованные мероприятия, увеличился незначительно. Атаки китайских хакеров на системы Минобороны оказались достаточно эффективными.

Китайская народная республика. Китай уже давно включил термин «информационная война» в лексикон своих военных специалистов. Сегодня он неуклонно движется к формированию единой доктрины информационной войны. Фактически, если революция в военном деле определяется как существенное изменение в технологии, дающее преимущество в военном обучении, организации, стратегии и тактике военных действий, то, возможно, Китай из всех стран сегодня испытывает истинную революцию в киберпространстве. Китайская концепция информационной войны включает уникальные китайские представления о войне вообще, основанные на современной концепции «народной войны», 36 стратагемах великого Сун Цзы, а также местных представлениях о том, как воевать на стратегическом, оперативном и тактическом уровне. Многие из его подхода имеет отношение к акценту на обмане, войне знаний и поиске асимметричных преимуществ над противником. Информационная война определена как «переход от механизированной войны индустриального возраста к войне решений и стиля управления, войне за знания и войне интеллекта». Китай развивает концепцию Сетевых сил (воинские подразделения численностью до батальона), которые состояли бы из высококлассных компьютерных экспертов, обученных в множестве государственных университетов, академий и учебных центров. Основной акцент делается на привлечение активной молодежи.

Великобритания. Британское представление об информационной войне схоже с американским. Это определение информационной войны как действий, воздействующих на информационные системы противника, при одновременной защите собственных систем. Кроме того, Великобритания использует юридическую структуру, основанную на существующих законах, которая в значительной степени может применяться к действиям в киберпространстве – Regulation of Investigatory Powers Act (RIP), принятый в 2000 г. Он предлагает, что нападения на информационные системы может рассматриваться как обычное уголовное преступление со всеми вытекающими последствиями. Данный акт позволяет британскому правительству перехватывать и читать электронную почту, а также требовать расшифровки личных файлов по требованию государственных чиновников.

Франция. Французы рассматривают концепцию информационной войны, состоящей из двух главных элементов: военной и экономической (или гражданской). Военная концепция предполагает несколько ограниченную роль информационных операций и направлена на миротворческие цели. В этом контексте союзники не рассматриваются как противники. Напротив, экономическая или гражданская концепция включает более широкий диапазон потенциального применения информационных операций. В этом случае французы не видят себя связанными рамками НАТО, ООН или согласием США. Их подход к экономическому конфликту позволяет быть и союзником, и противником одновременно. Французы даже имеют экономическую школу для информационной войны. Франция активно формирует структуры по контролю ее граждан в киберпространстве. Есть информация о том, что французы создали собственную версию систе-

мы «Эшелон» (по сообщениям американской прессы, система направлена на перехват фактически всех частных глобальных коммуникаций). Frenchelon, так некоторые назвали эту систему, по сообщениям, используется для контроля и анализа французских коммуникаций, особенно в районе Парижа [5].

На сегодняшний день Соединенные Штаты Америки обладают половиной всего мирового информационного ресурса. Ученым и военным этой страны удалось сбить вышедший из строя спутник, находившийся на орбите. Данный факт говорит об успешной работе над программой «Звездных войн», подписанной Бушем в ответ на попытку Пекина ослепить с помощью мощного лазера американский спутник. Россия и Китай внесли на рассмотрение конференции ООН по разоружению проект договора «О недопущении размещения вооружений в открытом космосе», который дополнял 40-летний рамочный договор, провозглашающий мирное освоение космоса. Ранее инициативы подобного рода были произведены в 1997 г. Госдумой, а затем Межпарламентской ассамблеи СНГ, которые обратились в ООН, ОБСЕ и Совет Европы с предложением о принятии международной конвенции о запрещении информационных войн и ограничении оборота информационного оружия. В марте 1998 г. этот вопрос был поднят на встрече с генсеком ООН Кофи Аннано и по инициативе России включен в повестку Генеральной ассамблеи ООН. Но без согласия Вашингтона любые инициативы в этой области останутся пустыми словами. Поэтому информационное оружие или, иначе, средство ведения бесконтактной войны шестого поколения – единственное не попадающее не под один договор оружие, что является недопустимым.

Литература

1. Военная разведка [Электронный ресурс]. – Режим доступа: <http://www.vrazvedka.ru/main/analytical/lekt-03.shtml>, своб.
2. Концепция национальной безопасности Российской Федерации. – М., 2000.
3. Советская военная энциклопедия. – М., 1978. – Т. 6.
4. Libicki M. What is Information Warfare. – National Defense University. ACIS paper 3. – August 1995.
5. Новости NEWSru.com, «Россия приступила к разработке «информационного оружия» против врага. Мир готовится к глобальным информационным войнам» от 31 января 2008 г. [Электронный ресурс]. – Режим доступа: <http://www.newsru.com/russia/31jan2008/cyberwar.html>, своб.
6. Доктрина информационной безопасности Российской Федерации. – М., 2000.

МЕТОДИКА ПОСТРОЕНИЯ МОДЕЛИ УГРОЗЫ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Н.Ю. Дрюков, И.В. Ермаков, Н.В. Ермаков
Научный руководитель – д.т.н., профессор Ю.А. Гатчин

В работе рассматриваются вопросы моделирования угроз информационной безопасности, основывающиеся на определении стоимости защищаемой информации для нарушителя и оценки эффективности применяемых защитных механизмов.

Введение

В условиях бурно развивающейся борьбы за информационные ресурсы и становления информационного общества крайне остро стоит вопрос моделирования и оценки угроз безопасности информации. Для любой информационной системы весь окружающий мир представляет собой одну сплошную угрозу. Атаки бывают как совершенно любительскими, совершаемыми в большинстве своем из хулиганских побуждений, так и профессиональными, проводимыми хорошо подготовленными специалистами, направленными на достижение конкретной цели. И те, и другие могут увенчаться успехом, однако в большинстве случаев атаки второй категории наносят намного больший урон. Ситуация осложняется тем фактом, что количество угроз растет в геометрической прогрессии, и поэтому производители оборудования и программного обеспечения и администраторы безопасности не всегда могут успеть адекватно среагировать на новую опасность. Кроме того, сам факт появления новой уязвимости до распространения информации о ней и подготовки рекомендаций и исправлений производителем ставит под удар все подверженные ей системы. Таким образом, защищаемая сторона всегда оказывается «догоняющей».

В качестве возможного решения данной проблемы может выступить моделирование и оценка опасности угроз. Опорной точкой для решения данной задачи является мотивирование и стратегия атакующего, так как именно он является зачинателем конфликта. Защищаемая же сторона вынуждена готовить адекватный ответ на применяемые атаки.

Угрозы безопасности информации

Источники угрозы информационной безопасности можно разделить на три группы:

1. естественные;
2. искусственные непреднамеренные;
3. искусственные преднамеренные.

Если рассмотреть источники, не связанные с человеческим фактором, т.е. относящиеся к первым двум группам, то модель нарушения информационной безопасности будет сводиться к реализации сбоя системы, причем вне зависимости от фактического нарушения. Иными словами, построение модели данной угрозы производится на основе моделирования нарушения информационной безопасности исключительно при рассмотрении информации как объекта защиты, а не объекта нападения.

Третья группа источников угроз представляет наибольший интерес для исследования. Реализация данных угроз приводит не только к урону для собственника информации, но и получению преимуществ для нарушителя. Строго говоря, целесообразность защиты информации и попыток ее обойти можно определить за счет рассмотрения стоимости информации с точки зрения нарушителя и стоимости информации с точки зрения собственника.

Для определения первого параметра необходимо рассмотреть мотивы нарушителя, а именно конечную выгоду, получаемую непосредственным нарушителем, при атаке на информационную безопасность.

В качестве исходных данных для исследования будут выступать: множество защищаемой информации $U = \{u_1, u_2, \dots, u_n\}$, множество возможных на нее воздействий – нарушений информационной безопасности $H = \{h_1, h_2, \dots, h_n\}$ и возможные мотивы нарушителя. Элементами множества нарушений информационной безопасности (НИБ) будут являться:

- 1) нарушение конфиденциальности
 - а) ограниченному числу лиц,
 - б) неограниченному числу лиц;
- 2) нарушение целостности
 - а) модификация,
 - б) уничтожение;
- 3) нарушение доступности
 - а) временное,
 - б) постоянное.

Схема нарушения информационной безопасности представляет собой взаимосвязь элементов трех множеств – мотив, нарушение, информация.

Анализ мотивов нарушителя

Для проведения анализа мотивов нарушителя следует принять за истину три следующих тезиса.

- 1) Вся информация находится в незащищенном состоянии, т.е. внешний мир может осуществлять над информацией любые действия.
- 2) Мотивы злоумышленника:
 - борьба с конкурентами;
 - получение закрытой информации;
 - вымогательство.
- 3) Предпринимаемые действия увенчиваются успехом и приводят к реализации мотивов, существующих вне времени.

Другими словами, выделяемая информация представляет ценность для злоумышленника в любой момент времени, что в реальном мире не является истиной. Это вызвано тем, что стоимость информации определяется контекстом использования, зависящим от модели окружающего мира в конкретный момент времени, но, для упрощения анализа и исключения ошибочной оценки, из контекста убирается текущее время.

Исключив защиту, можно определить соответствие между мотивами и искомой информацией. Вывод достаточно очевиден. Любая информация, отсутствующая в открытых источниках, становится объектом второго мотива. Любая информация, от которой зависит деятельность организации, неприменимая для злоумышленника и критичная для собственника, становится объектом 3-го мотива. Информация, являющаяся объектом 3-го мотива, а также информация, применимая для злоумышленника, становится объектом 1-го мотива. Для более тщательного описания данной связи необходимо анализировать действия нарушителя после успешной реализации атаки, т.е. реальные последствия данного воздействия. Анализ осуществляется на основе информационных процессов в организации, причем не играет роли, как они проходят, важность представляет только содержимое, а также внешнее окружение организации.

Связующим звеном между информацией и мотивом является НИБ. С одной стороны, оно может быть неприменимо к критичной информации, а с другой – оно может быть неприменимо к мотиву.

Применимость НИБ к информации.

Для пары элементов «НИБ» и «информация» есть свое значение вероятности. Возможность нарушения определяется наличием у информации уязвимости, которая может породить данное нарушение. Таким образом, искомая вероятность будет рассчитана по формуле:

$$p(n_j) = \sum_{i=1}^n p(y_i) * p(n_j/y_i). \quad (1)$$

Вероятность $p(y_i)$ основывается на воздействии неконтролируемых факторов, таких как человеческий фактор, сбои электропитания, несчастные случаи и стихийные бедствия и т. д. Вероятность появления данных уязвимостей рассчитывается на основе статистических данных. Остальные вероятности определяются надежностью применяемых технических средств и равняются отношению количества неуспешных атак к их общему числу. Работы, посвященные исследованию эффективности отдельных средств защиты, проводимые различными независимыми организациями, несложно найти в открытых источниках.

Вероятность $p(n_j/y_i)$ основывается на возможности наступления нарушений информационной безопасности при реализации данной уязвимости. Значение равно 0, если уязвимость и нарушение несовместимы, в остальных случаях это значение определяется эффективностью защитных механизмов всей системы, ее комплексности, т.е. насколько эффективно вся система может противостоять «открытым» уязвимостям.

В результате получится список критичных уязвимостей, которые могут привести к нарушению информационной безопасности. Ими будут являться уязвимости, у которых хотя бы одно из значений $p(n_j/y_i)$ будет отличным от 0.

Применимость НИБ к мотиву

Злоумышленник, опираясь на собственные мотивы, выбирает наиболее подходящие возможные НИБ для этого. Очевидно, что не каждое из возможных нарушений соответствует мотивам нарушителя. Достижение какого-либо НИБ без привязки к информации удовлетворит злоумышленника с какой-либо вероятностью $p(m_i/n_j)$.

Соответственно, мотивы и НИБ также могут быть совместимы (когда достижение рассматриваемого НИБ однозначно удовлетворяет конкретному мотиву нарушителя, и вероятность $p(m_i/n_j)$ будет равна 1), ограниченно совместимыми (когда достижение рассматриваемого НИБ удовлетворит конкретному мотиву с какой-либо вероятностью $p(m_i/n_j)$, находящейся между 0 и 1) и несовместимыми (когда достижение рассматриваемого НИБ не при каких обстоятельствах не приведет к удовлетворению мотива, и вероятность $p(m_i/n_j)$ будет равна 0). Совместимость мотивов и НИБ приведена в табл. 1.

Таким образом, удовлетворение мотивов нарушителя возможно при реализации конкретных нарушений для конкретной информации, при условии, что $p(m_i/n_j)$ больше 0 и мотив нацелен на данную информацию.

Вероятность удовлетворения мотива нарушителя при реализации критичных уязвимостей по каждой отдельной информации будет рассчитываться по формуле:

$$p(m) = \sum_{j=1}^n \sum_{i=1}^m p(y_i) * p(n_j/y_i) * p(m/n_j). \quad (2)$$

Определение вероятности реализации всего мотива по всему множеству объектов информации является тривиальной задачей, определяемой механизмом реализации мотива, совместимостью НИБ. В общем и целом это рассчитывается при помощи обычных законов теории вероятности сложением и перемножением вероятностей.

Мотивы НИБ	Борьба с конкурентами	Получение закрытой информации	Вымогательство
Нарушение конфиденциальности ограниченному числу лиц	$p(m/n_j)$	1	0
Нарушение конфиденциальности неограниченному числу лиц	$p(m/n_j)$	$p(m/n_j)$	1
Нарушение целостности: модификация	$p(m/n_j)$	0	1
Нарушение целостности: уничтожение	$p(m/n_j)$	0	1
Нарушение доступности: временное	$p(m/n_j)$	0	1
Нарушение доступности: постоянное	$p(m/n_j)$	0	1

Таблица 1. Совместимость мотивов нарушителя и нарушений информационной безопасности

Для удобства рассмотрения поведения нарушителя предположим, что он преследует единственный мотив, который в той или иной степени с определенной вероятностью $p(m/n_j)$ достигается при реализации любого из применимых нарушений. Величина $p(m)$ из формулы (2), кроме $p(m/n_j)$, также зависит от величин $p(y_i)$ и $p(n_j/y_i)$. Данные величины не зависят от действия нарушителя и определяются защитными механизмами. Исходя из этого, можно построить матрицу вариантов направлений атак, элементами в которой будут являться пары «нарушение»–«информация» ($u_i n_j$), применимые для конкретного мотива, для которых $p(m/n_j)$ больше 0. Данная матрица представлена в табл. 2.

Варианты защитных стратегий Возможные цели	Варианты направлений атак						
	A_1	A_2	A_3	A_4	A_5	...	A_k
$(u_1 n_1)$	c_{11}^1	c_{11}^2	c_{11}^3	c_{11}^4	c_{11}^5	...	c_{11}^k
$(u_1 n_2)$	c_{12}^1	c_{12}^2	c_{12}^3	c_{12}^4	c_{12}^5	...	c_{12}^k
$(u_1 n_3)$	c_{13}^1	c_{13}^2	c_{13}^3	c_{13}^4	c_{13}^5	...	c_{13}^k
$(u_2 n_1)$	c_{21}^1	c_{21}^2	c_{21}^3	c_{21}^4	c_{21}^5	...	c_{21}^k
$(u_2 n_2)$	c_{22}^1	c_{22}^2	c_{22}^3	c_{22}^4	c_{22}^5	...	c_{22}^k
...
$(u_i n_j)$	c_{ij}^1	c_{ij}^2	c_{ij}^3	c_{ij}^4	c_{ij}^5	...	c_{ij}^k

Таблица 2. Варианты направлений атак

Соответственно, для каждой из пар «нарушение»–«информация» существуют различные способы обеспечения безопасности с разной эффективностью. Комбинация данных средств обозначена как A_k . Элемент таблицы c_{ij}^k можно определить по формуле:

$$c_{ij}^k = p(y_k) * p(n_j/y_k) * p(m/n_j) * C_{u_i}, \quad (3)$$

где C_{u_i} – ценность информации с точки зрения нарушителя.

Согласно математической теории игр, оптимальной для нарушителя будет «максиминная стратегия». Ориентируясь на разумные действия противника (защищающейся стороны), он делает вывод, что был выбран вариант защиты, минимизирующий выигрыш от нарушения, $c_{ij} = \min_k c_{ij}^k$. Полученное подмножество значений c_{ij} определяет минимально возможные выигрыши вне зависимости от выбора атаки. Если из полученных значений выбрать максимальное, $c = \max_{ij} c_{ij}$, то получится оптимальное направление, гарантирующее выигрыш от атаки не меньше c , которое будет характеризоваться набором $u_i n_j$, где i и j удовлетворяют следующему равенству:

$$c = \max_{ij} \min_k c_{ij}^k = \max_{ij} \min_k (p(y_k) * p(n_j / y_k) * p(m / n_j) * C_{u_i}).$$

Для обороняющейся стороны оптимальной будет «минимаксная стратегия». Ориентируясь на разумные действия противника (нарушителя), он делает вывод, что был выбран вариант атаки, максимизирующий выигрыш от нарушения, $c^k = \max_{ij} c_{ij}^k$. Полученное подмножество значений c^k определяет максимально возможные выигрыши вне зависимости от выбора защиты. Если из полученных значений выбрать минимальное, $c = \min_k c^k$, то получится оптимальное направление защиты, гарантирующее выигрыш от атаки не больше c , которое будет характеризоваться набором A_k , где k удовлетворяют следующему равенству:

$$c = \min_k \max_{ij} c_{ij}^k = \min_k \max_{ij} (p(y_k) * p(n_j / y_k) * p(m / n_j) * C_{u_i}).$$

Эффективность подобного выбора стратегий напрямую зависит от секретности применяемых стратегий, поэтому применение данного подхода является наиболее эффективным при проектировании защищенных систем.

Предложенная в данной работе методика, базирующаяся на оценке стоимости информации, теории вероятности и теории игр, позволяют осуществить моделирование поведения нарушителя согласно его заинтересованности в защищаемой информации. Данный подход позволяет оптимизировать и оценивать эффективность всей системы защиты по различию важности данных, обрабатываемых в защищаемых системах.

Литература

1. Вентцель Е.С. Элементы теории игр. – М.: Государственное издательство физико-математической литературы, 1961. – 72 с.
2. Домарев В.В. Безопасность информационных технологий. Методология создания систем защиты. – Киев: ТИД «ДС», 2001. – 688 с.
3. Мельников В.П., Клейменов С.А., Петраков А.М. Информационная безопасность и защита информации. – М.: Academia, 2007. – 336 с.
4. Палий И.А. Введение в теорию вероятностей. – М.: Высшая школа, 2005. – 147 с.
5. Шумский А.А., Шелупанов А.А. Системный анализ в защите информации. – М.: Гелиос АРВ, 2005. – 224 с.

**Научная школа
«НАНОМАТЕРИАЛЫ, НАНОСТРУКТУРЫ,
НАНОТЕХНОЛОГИИ»**

СПЕКТРАЛЬНО-ЛЮМИНЕСЦЕНТНОЕ ИССЛЕДОВАНИЕ ВЗАИМОДЕЙСТВИЯ КВАНТОВЫХ ТОЧЕК С ТЕТРАПИРРОЛЬНЫМИ МОЛЕКУЛАМИ

А.А. Степанов, А.О. Орлова

Научный руководитель – д.ф.-м.н., вед.н.с. В.Г. Маслов

Проведено спектрально-люминесцентное исследование взаимодействия заряженных квантовых точек (КТ) с тетрапиррольными молекулами, относящимися к классу порфиринов. Показано, что КТ в водном растворе при взаимодействии с противоположно заряженными молекулами образуют ассоциаты, в которых вокруг одной молекулы порфирина может образовываться конгломерат из 6, 7 и более квантовых точек. При этом происходит эффективное тушение как люминесценции КТ, находящихся в образующемся конгломерате, так и порфириновой молекулы, вокруг которой этот конгломерат образуется. Обнаружено наличие нескольких спектрально различных форм порфирина, находящихся в состоянии ассоциации с КТ. Предложена математическая модель тушения люминесценции КТ и порфириновых молекул в зависимости от соотношения их концентраций. Обнаружены спектрально-люминесцентные проявления переноса энергии от КТ к молекулам порфирина, находящимся в состоянии ассоциации с КТ.

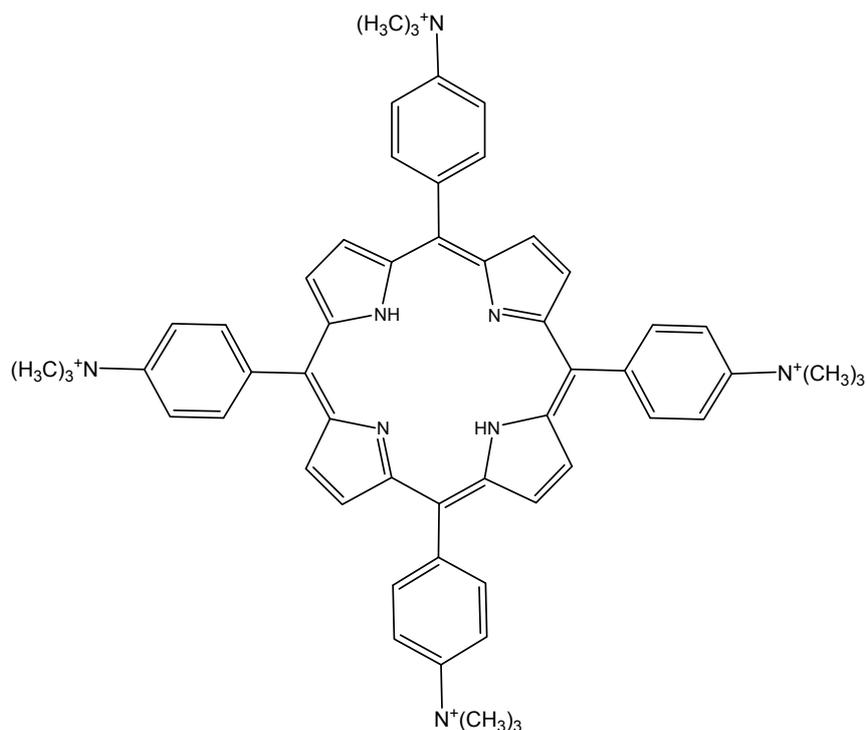
Введение

Взаимодействие полупроводниковых квантовых точек (КТ) с органическими молекулами ароматических соединений или красителей активно исследуются в последнее время, в частности, в связи с задачами создания сенсibilизаторов для фотодинамической терапии рака и химических сенсоров. Особый интерес представляют системы, в которых возможен безызлучательный перенос энергии (FRET, Fluorescence Resonance Energy Transfer) между КТ и молекулами, поскольку этот механизм может лежать в основе функционирования сенсоров. Ранее перенос энергии в ассоциатах КТ разного размера подробно исследовался в работе [1]. В системах типа КТ/органическая молекула наиболее эффективного переноса энергии между КТ и молекулой можно ожидать в случае непосредственного контакта между ними. Однако имеется только лишь небольшое число работ [2, 3], в которых такой непосредственный контакт имеется, и во всех этих случаях указывалось на наличие некоторых механизмов взаимодействия, не сводящихся к FRET. Отсюда вытекает необходимость постановки данной работы, в которой одной из задач является исследование спектрально-люминесцентных проявлений взаимодействия КТ с органическими молекулами в растворе в условиях, когда можно ожидать непосредственного контакта между ними, в силу электростатического типа взаимодействия. Следует также иметь в виду, что электростатический механизм рассматривается как один из важнейших механизмов создания супрамолекулярных систем. В работе [1] было продемонстрировано образование ассоциатов, состоящих из разноименно заряженных КТ разного размера. Было продемонстрировано, что в таких ассоциатах наблюдаются ярко выраженные спектрально-люминесцентные проявления однонаправленного переноса энергии между КТ. В данной работе делается попытка построения подобного рода ассоциатов, состоящих из разноименно заряженных КТ и органической молекулы.

Объекты исследования

В работе использовались отрицательно заряженные водорастворимые CdTe квантовые точки с максимумом длинноволновой полосы поглощения в области от 515 до 535 нм и положениями соответствующих максимумов люминесценции в области 540–560 нм.

В качестве органической молекулы был использован безметалльный тетра(*p*-триметиламино)фенилпорфин (ТАП):



Спектры поглощения и люминесценции ТАП в водном растворе приведены на рис. 1, а, б.

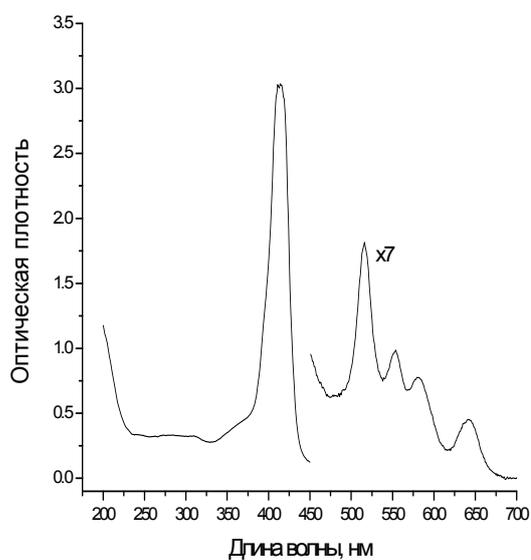


Рис. 1, а.

Спектр поглощения водного раствора ТАП

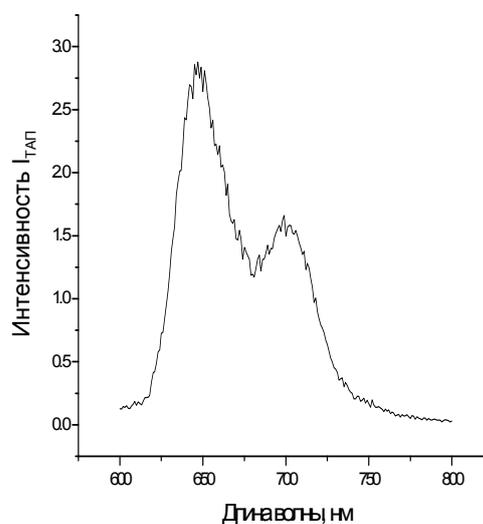


Рис. 1, б.

Спектр люминесценции водного раствора ТАП при возбуждении 582 нм

Поскольку, как указано выше, в данной работе применялись только КТ с положением максимума полосы люминесценции в области 540–560 нм, между спектром люминесценции КТ и спектром поглощения ТАП по всем случаям имело место вполне заметное перекрытие, что делает принципиально возможным перенос энергии от КТ на ТАП. Перенос же энергии от ТАП на КТ оказывается практически невозможным из-за отсутствия перекрытия между спектрами люминесценции ТАП и поглощения КТ.

Условия проведения эксперимента и полученные результаты

На рис. 2 приведены спектры поглощения и люминесценции одного из типов КТ используемых в данной работе. Эти КТ имеют диаметр 2,5 нм, коэффициент экстинкции в длинноволновой полосе поглощения составляет $\varepsilon=6,8 \cdot 10^4$.

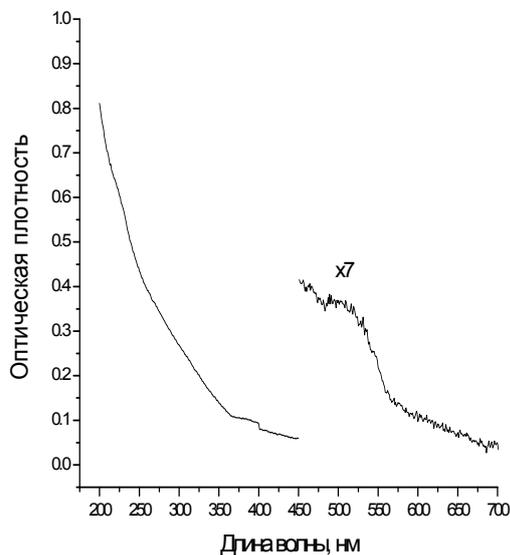


Рис. 2, а.
Спектр поглощения водного раствора КТ

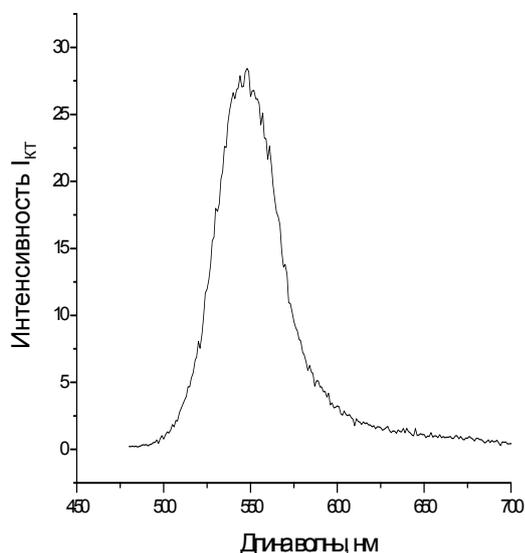


Рис. 2, б. Спектр люминесценции водного раствора КТ при возбуждении 460 нм

На рис. 3 показаны изменения спектра люминесценции КТ при добавлении последовательных порций ТАП. Осуществлялось последовательное добавление к 2 мл раствора КТ концентрации $3,1 \cdot 10^{-7}$ М одинаковых порций раствора ТАП объема 17 мкл и концентрации $7,1 \cdot 10^{-6}$ М. Как видно из рис. 3, наблюдалось исключительно эффективное тушение люминесценции КТ.

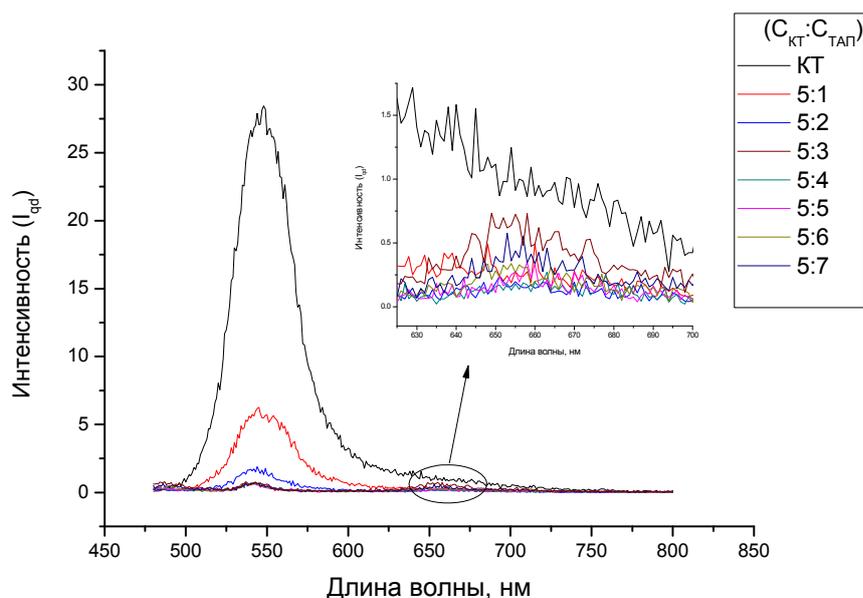


Рис. 3. Спектры люминесценции водных растворов КТ для различных относительных концентраций донора и акцептора

Существенно, что при добавлении первых порций ТАП не наблюдалось никакой люминесценции ТАП. Однако, начиная с некоторого значения концентрации ТАП, его люминесценция появляется и возрастает по мере добавления.

На рис. 4 показана построенная по данным, приведенным на рис. 3, зависимость относительной интенсивности люминесценции КТ (I/I_0 , где I и I_0 – интенсивности люминесценции КТ в присутствии ТАП и в его отсутствии, соответственно) от относительной концентрации ТАП ($x=C_p/C_0$, где C_0 и C_p – концентрация КТ и концентрация добавленного ТАП, соответственно). На рис. 4 также показана концентрационная зависимость относительной интенсивности люминесценции ТАП (I_p/I_{p0} , где I_p – измеренная интенсивность люминесценции ТАП, I_{p0} – интенсивность люминесценции ТАП в отсутствии КТ).

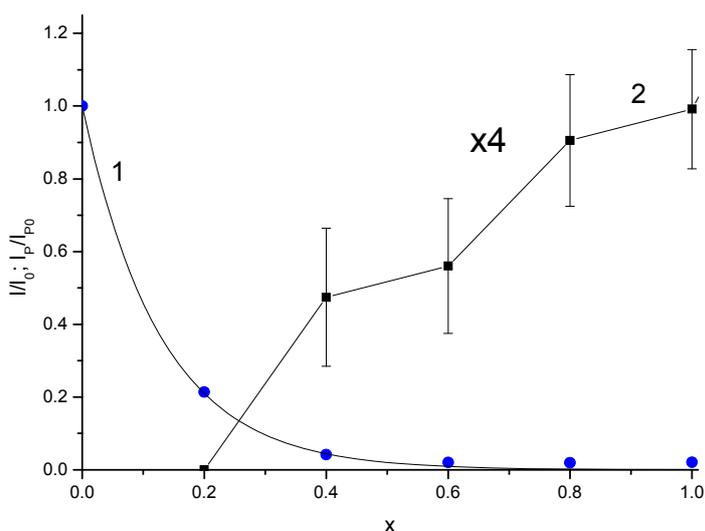


Рис. 4. Зависимость относительной интенсивности люминесценции КТ (1) и относительной интенсивности люминесценции добавляемого ТАП (2) от величины $x=C_p/C_0$.

Сплошная кривая – аппроксимация экспонентой $\frac{I}{I_0} = e^{-7.8x}$

Как показал анализ зависимости I/I_0 от x , она близка к экспоненциальной: $\frac{I}{I_0} = e^{-nx}$

с показателем экспоненты n около 8. Это означает, что при действии первых малых порций ТАП одна молекула ТАП тушит люминесценцию не менее, чем 7 КТ. Если следовать статической модели тушения, то по аналогии с работой [2] следует полагать, что появление в растворе молекулы ТАП приводит к образованию вокруг нее конгломерата из n КТ. При этом люминесценция всех этих n КТ полностью потушена. Существенно, что люминесценция молекулы ТАП, оказавшейся в центре такого конгломерата, оказывается также полностью потушенной. Заметим, что тушение люминесценции КТ молекулой ТАП может осуществляться по механизму безызлучательного переноса энергии, в то время как от ТАП на КТ невозможен, в силу взаимного расположения спектров люминесценции ТАП и поглощения КТ. В этом случае тушение люминесценции ТАП может быть обусловлено другими механизмами, например, переносом электрона.

Что же касается зависимости I_p/I_{p0} от x , то после достижения некоторой величины относительной концентрации ТАП интенсивность люминесценции ТАП растет приблизительно линейно с ростом относительной концентрации ТАП. При этом оказывается, что интенсивность люминесценции ТАП оказывается в 4–5 раз ниже, чем она была бы в отсутствии КТ.

Тот факт, что интенсивность люминесценции добавляемого ТАП оказывается существенно ниже, чем у свободного ТАП в отсутствие КТ, наводит на мысль о том, что эта люминесценция обусловлена какой-то связанной с КТ формой ТАП, хотя, возможно, и слабо связанной, поскольку приведенные на рис. 3 спектры люминесценции в области полос ТАП существенно не отличаются от спектров люминесценции свободного ТАП (см. рис. 1, б). Чтобы выяснить более подробно спектральные свойства как люминесцирующих, так и, возможно, нелюминесцирующих форм ТАП, появляющихся по мере добавления ТАП к раствору КТ, было проведено более подробное исследование спектров поглощения и возбуждения люминесценции при добавлении ТАП к раствору КТ при несколько больших концентрациях КТ, чем описано выше. В этих экспериментах использовались КТ, спектры поглощения и люминесценции которых приведены на рис. 5. Эти КТ характеризуются диаметром 2,75 нм и коэффициентом экстинкции в длинноволновой полосе поглощения $\epsilon=8,55 \cdot 10^4$.

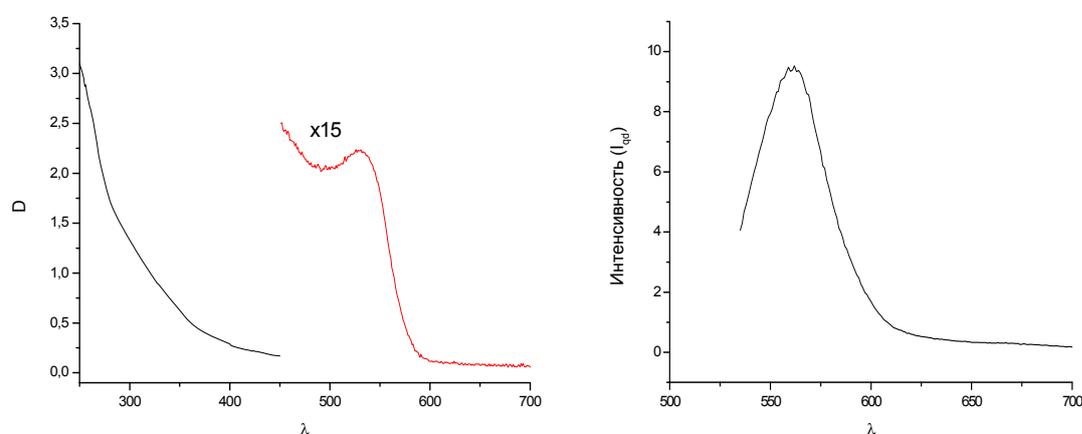


Рис. 5. Спектры поглощения (а) и люминесценции при возбуждении светом 515 нм (б) водных растворов КТ, используемых в экспериментах

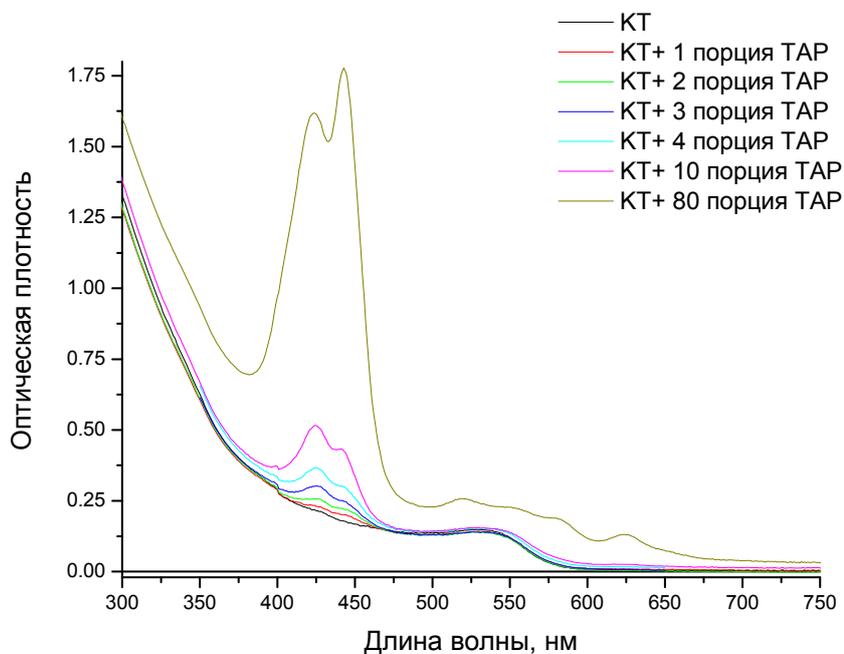


Рис. 6. Спектры поглощения водных растворов КТ после последовательных добавлений указанного числа порций ТАП

Спектры поглощения при добавлении последовательных порций раствора ТАП концентрации $8,3 \cdot 10^{-5}$ М и объема 1,7 мкл к раствору КТ концентрации $1,7 \cdot 10^{-6}$ М приведены на рис. 6. На рис. 7 приведены спектры возбуждения люминесценции, измеренные при регистрации на нескольких длинах волн в области максимума люминесценции ТАП.

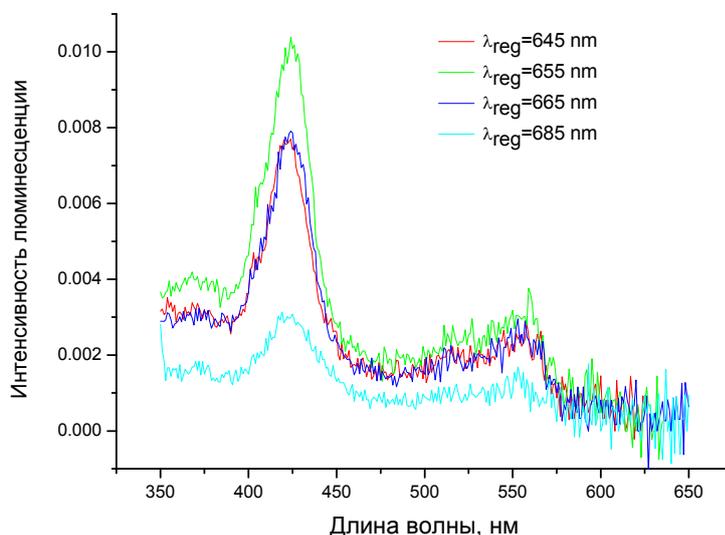


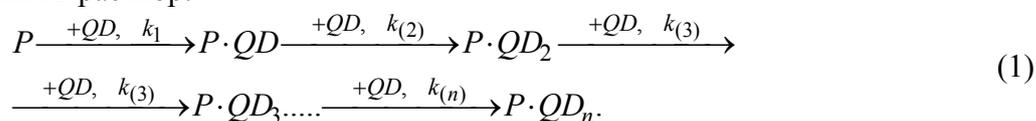
Рис. 7. Спектры возбуждения люминесценции водного раствора КТ после добавления 10 порций ТАП (см. рис. 6) при регистрации в области полосы люминесценции ТАП

Сопоставление приведенных на рис. 6 и 7 спектров показывает, что наблюдаемой в смесях растворов КТ с ТАП люминесцирующей форме ТАП с полосой люминесценции у 650 нм соответствует интенсивная полоса в спектрах поглощения и возбуждения 425 нм. Такое положение данной полосы существенно отличается от ее положения у свободного ТАП в отсутствие КТ (415 нм, см. рис. 1). Это обстоятельство согласуется с представлением об этой форме как о форме ТАП, ассоциированной с КТ, хотя и сравнительно слабо. Однако, кроме этой формы, в спектре поглощения наблюдается появление полосы с максимумом 443 нм, аналог которой отсутствует в спектре возбуждения люминесценции. Очевидно, эта полоса принадлежит другой, нелюминесцирующей (или очень слабо люминесцирующей) форме ТАП.

Анализ приведенных на рис. 7 спектров указывает на наличие переноса энергии от КТ к ТАП. Об этом может свидетельствовать подъем спектральной кривой в коротковолновую сторону на спектре, измеренном при регистрации на длине волны 655 нм, т.е. вблизи максимума люминесценции данной формы ТАП, если эту кривую сопоставить со спектрами, измеренными при регистрации на 645 и 665 нм. Следует заметить, однако, что этот подъем довольно слабо выражен, что может свидетельствовать о малой эффективности переноса энергии, например, из-за конкурирующих процессов тушения возбуждения КТ.

Анализ полученных результатов, модель тушения люминесценции

Согласно развиваемым здесь представлениям, тушение люминесценции обусловлено образованием конгломератов КТ вокруг молекул ТАП (P) при добавлении последних в раствор:



Если бы отсутствовали какие-либо иные процессы, связывающие молекулы ТАП, то, как легко видеть, это приводило бы к строго линейной зависимости относительной концентрации свободных КТ $\frac{C_{QD}}{C_{QD}^0}$ от x :

$$\frac{C_{QD}}{C_{QD}^0} = 1 - nx. \quad (2)$$

Мы здесь предполагаем, что люминесцируют только свободные КТ, т.е. $\frac{I}{I_0} = \frac{C_{QD}}{C_{QD}^0}$. При этом зависимость от x относительной интенсивности люминесценции ТАП $\frac{I_P}{I_{P0}}$, очевидно, должна иметь следующий вид:

$$\frac{I_P}{I_{P0}} = \begin{cases} 0 & \text{при } 0 \leq x < \frac{1}{n} \\ x - \frac{1}{n} & \text{при } x \geq \frac{1}{n} \end{cases}. \quad (3)$$

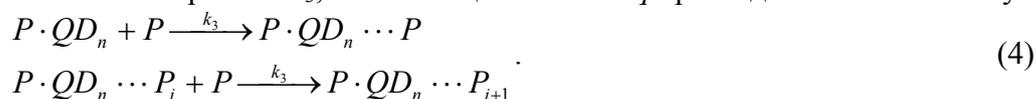
Это означает, что люминесценция ТАП появляется только после исчерпания всех свободных КТ, и ее квантовый выход соответствует свободным молекулам ТАП, на что указывает коэффициент при x , равный 1.

Полученные экспериментальные результаты не согласуются с такой моделью: зависимость $\frac{C_{QD}}{C_{QD}^0}$ от x на самом деле имеет вид, близкий к экспоненциальному. Зависимость же

$\frac{I_P}{I_{P0}}$ от x , хотя и близка к линейной, но с коэффициентом, существенно меньшим единицы.

Это требует дополнительных допущений в рамках рассматриваемой модели. Так, можно предполагать, что если в растворе уже присутствуют конгломераты $P \cdot QD_n$, то при добавлении ТАП в раствор может происходить образование более слабых ассоциатов $P \cdot QD_n \cdots P_i$ со скоростью, конкурирующей с образованием конгломератов КТ вокруг ТАП. Будем также предполагать, что люминесценция таких слабо ассоциированных молекул ТАП может быть потушена не полностью, а в 4–6 раз по сравнению со свободными молекулами ТАП. Предполагается также, что число q может быть велико, т.е. в растворе при не может быть свободных молекул ТАП (при разумных значениях x).

Будем для простоты предполагать, что процесс образования таких ассоциатов идет с одной и той же скоростью k_3 , не зависящей от числа q присоединившихся молекул ТАП:



Обозначим через C_c текущую суммарную концентрация всех конгломератов типа $P \cdot QD_n \cdots P_i$ ($i \geq 0$) и через C_p – концентрацию добавляемого ТАП. Тогда, поскольку скорости вовлечения молекул ТАП в реакции определяется скоростями k_1 и k_3 , добавление малой концентрации dC_p молекул ТАП приведет к убыли концентрации свободных КТ, определяемой следующим выражением:

$$dC_{QD} = -dC_p \frac{k_1 C_{QD}}{k_1 C_{QD} + k_3 C_c} \quad (5)$$

Поскольку число КТ в том или ином состоянии не меняется и равно исходной концентрации КТ C_{QD}^0 , имеем: $C_{QD}^0 = C_{QD} + nC_C$, или: $C_C = \frac{C_{QD}^0 - C_{QD}}{n}$. Отсюда получаем уравнение

$$\frac{dC_{QD}}{dC_P} = -n \frac{k_1 C_{QD}}{k_1 C_{QD} + \frac{k_3}{n} C_{QD}^0 + \frac{k_3}{n} C_{QD}}. \quad (6)$$

Если ввести обозначения

$$y = \frac{C_{QD}}{C_{QD}^0}, \quad x = \frac{C_P}{C_{QD}^0}, \quad b = \frac{k_3}{nk_1}, \quad a = 1 - b, \quad (7)$$

то получаем уравнение

$$\frac{dy}{dx} = -n \frac{1}{a + \frac{b}{y}}, \quad (8)$$

или $\frac{dx}{dy} = -\frac{1}{n} \left(a + \frac{b}{y} \right)$ с начальным условием: $y|_{x=0} = 1$ (9)

Интегрируя, получаем:

$$x = -\frac{a}{n} y - \frac{b}{n} \ln |y| + C \quad (10)$$

Величина C определяется из начального условия: $0 = \frac{a}{n} + C$, т.е. $C = -\frac{a}{n}$. Отсюда решение уравнения с начальным условием получается в следующем виде:

$$x = -\frac{a}{n}(y-1) - \frac{b}{n} \ln y, \quad \text{или} \quad x = -\frac{(1-b)}{n}(y-1) - \frac{b}{n} \ln y. \quad (11)$$

Это уравнение в общем случае не разрешается в явном виде относительно y . В частном случае, когда $b=1$ и $a=0$, имеем: $y = e^{-nx}$. В другом предельном случае, $b=0$ и $a=1$, имеем $y = 1 - nx$. Таким образом, данная модель может объяснить как экспоненциальную зависимость $\frac{I}{I_0}$ от x , так и возможные отклонения от такой зависимости, которые могут наблюдаться в некоторых случаях, но пока не были нами подробно проанализированы.

Рассмотрим теперь люминесценцию ТАП. Согласно сделанному выше предположению, люминесцируют только слабо ассоциированные молекулы ТАП (P), т.е. молекулы, входящие в состав комплексов $P \cdot QD_n \cdots P_q$. Очевидно, концентрация таких слабо ассоциированных молекул C_{PW} выражается как разность между числом добавленных молекул P и числом конгломератов: $C_{PW} = C_P - C_C$. Обозначим: $z = \frac{C_{PW}}{C_{QD}^0}$.

Поскольку $C_C = \frac{1-y(x)}{n}$, получаем:

$$z(x) = x - \frac{1-y(x)}{n}. \quad (12)$$

Поскольку $y \xrightarrow{x \rightarrow \infty} 0$, то у функции $z(x)$ имеется асимптота: $z = \left(x - \frac{1}{n} \right)$. Отсю-

да может быть получена кривая зависимости люминесценции ТАП от концентрации добавляемого ТАП. Если в качестве характеристики люминесценции ТАП принята

нормированная величина: I_p / I_p^0 , где I_p^0 – люминесценция свободного ТАП при концентрации ТАП, равной C_{QD}^0 , а интенсивность люминесценции молекул ТАП, находящихся в слабо ассоциированном состоянии, потушена по сравнению со свободными молекулами ТАП, в $1/\varphi$ раз, то отсюда получаем:

$$\frac{I_p}{I_p^0} = \varphi z(x). \quad (13)$$

Нетрудно убедиться, что такое поведение находится в полном согласии с изложенными экспериментальными данными.

Таким образом, в рамках данной модели «люминесцирующая» форма ТАП рассматривается как молекулы, слабо ассоциированные с конгломератом. В принципе в этом случае мог бы наблюдаться перенос энергии между КТ и этой формой ТАП, поскольку фёрстеровский радиус для такого переноса составляет десятки ангстрем. Полученные результаты не исключают возможность такого переноса, но свидетельствуют о том, что он не может иметь высокую эффективность. Это неудивительно, поскольку слабо ассоциированные молекулы ТАП присоединяются к уже не люминесцирующему конгломерату КТ, образовавшемуся вокруг значительно более тесно ассоциированной с конгломератом молекулы ТАП, которая служит эффективной ловушкой энергии, поглощаемой конгломератом. Такие тесно связанные молекулы ТАП, возможно, хотя бы частично обуславливают наличие описанной выше нелюминесцирующей форме ТАП (с максимумом поглощения 443 нм).

Основные результаты и выводы

1. Квантовые точки в водном растворе при взаимодействии с противоположно заряженными порфириновыми молекулами соединений образуют ассоциаты, в которых вокруг одной молекулы порфирина может образовываться конгломерат из 6, 7 и более квантовых точек. При этом происходит эффективное тушение как люминесценции КТ, находящихся в образующемся конгломерате, так и порфириновой молекулы, вокруг которой этот конгломерат образуется.
2. Обнаружено наличие нескольких спектрально различных форм порфирина, находящихся в состоянии ассоциации с КТ.
3. Обнаружены признаки переноса энергии от КТ к молекулам порфирина, находящимся в состоянии ассоциации с КТ.
4. Предложена математическая модель тушения люминесценции КТ и порфириновых молекул в зависимости от соотношения их концентраций.

Литература

1. R. Wargnier, A.V. Baranov, V.G. Maslov, V. Stsiapura, M. Artemyev, M. Pluot, A. Sukhanova, I. Nabiev. Energy transfer in aqueous solutions of oppositely charged CdSe/ZnS core/shell quantum dots and in quantum dot-nanogold assemblies // *Nano Letters*. – 2004. – V. 4. – № 3. – P. 451–457.
2. Орлова А.О., Маслов В.Г., Скалецкая И.Е., Баранов А.В. Перенос энергии в ассоциатах полупроводниковых квантовых точек с молекулами тетрапиридинпорфирина // *Оптика и спектроскопия*. – 2006. – 101(4). – С. 618–625.
3. Schmelz O., Mews A., Basche T., Herrmann A., Müllen K. Supramolecular complexes from CdSe nanocrystals and organic fluorophors // *Langumir*. – American Chemical Society. – 2001. – 17. – P. 2861–2865.

ЭФФЕКТ ФОТОННОЙ ЛАВИНЫ В НАНОСТРУКТУРАХ

Р.С. Левицкий

Научный руководитель – д.ф.-м.н., профессор Е.Ю. Перлин

Исследован процесс запуска фотонной лавины в твердотельных наноструктурах с различными типами электронной зонной структуры. Составлены балансные уравнения для концентраций носителей в подзонах размерного квантования. Решение систем нелинейных балансных уравнений для рассматриваемых структур показывает наличие характерных для эффекта фотонной лавины особенностей.

Введение

Взаимодействие интенсивного ИК излучения с полупроводниковыми структурами с глубокими квантовыми ямами при определенных условиях может привести к возникновению эффекта фотонной лавины. Впервые эффект фотонной лавины наблюдался в примесных системах редкоземельных ионов, и его использование стало одним из наиболее эффективных способов возбуждения коротковолновой люминесценции с помощью длинноволновой накачки [1]. За прошедшее время с момента обнаружения эффекта в 1979 г. был предложен и практически реализован целый ряд моделей лавинной апконверсии в примесных системах (подробную библиографию см. в [2]).

Эффект фотонной лавины характеризуется рядом отличительных свойств, к которым, в первую очередь, относятся: а) наличие четко выраженного порога – при интенсивности излучения накачки, равной пороговому значению (j_{th}), скачкообразным образом возрастает поглощение света накачки на переходах между возбужденными электронными состояниями и скачком возрастают заселенности этих состояний; б) время τ_{eq} установления квазиравновесного распределения электронов резко возрастает в области пороговых интенсивностей света j_{th} . Типичные времена τ_{eq} установления квазиравновесного распределения в электронной системе редкоземельных ионов при эффекте фотонной лавины составляют 1–100 мс при пороговых плотностях энергии накачки, необходимых для включения лавинного механизма, $E_{sw} \sim 0,1–10$ мкДж/мкм². Столь медленное протекание эффекта фотонной лавины в системах редкоземельных ионов, связанное с малыми значениями сил осцилляторов для актуальных оптических переходов, естественно, ограничивает круг возможностей практического использования этого явления в оптоэлектронике. В этой связи представляет интерес поиск твердотельных систем, где переключение системы осуществлялось бы за значительно более короткие времена с затратой меньшей энергии. Можно предположить, что подходящей системой окажется квантовая яма, где, с одной стороны, силы осцилляторов для переходов между подзонами размерного квантования являются величинами порядка единицы, а с другой стороны, достаточно коротки времена релаксации в электронной системе.

В настоящей работе исследован процесс запуска фотонной лавины в наноструктурах с различными типами электронной зонной структуры: (1) в материалах с легированными квантовыми ямами; (2) с квантовыми ямами типа II; (3) квантовыми ямами типа I. Для описания кинетики фотоиндуцированных переходов в системе составлялись балансные уравнения для концентраций носителей в подзонах размерного квантования квантовой ямы. Решения систем нелинейных дифференциальных уравнений получены с помощью численных методов.

Несмотря на одинаковую природу эффекта в различных системах, для каждой из них имеется ряд особенностей. Эти особенности могут играть существенную роль при практическом использовании эффекта фотонной лавины, в том числе для накачки коротковолновых лазеров длинноволновым светом, полностью оптического переключения и управления светом с помощью света.

1. Модели фотонной лавины для структур с квантовыми ямами

Для расчета динамики процесса типа фотонной лавины необходимо построить модель для каждой конкретной системы, т.е. выделить электронные или электрон-колебательные состояния, к квантовым переходам между которыми сводится кинетика лавины, а также выделить все существенные для эффекта фотонной лавины одно- и многочастичные переходы.

1.1. Легированная квантовая яма

Рассмотрим глубокую прямоугольную квантовую яму для электронов с шириной $2a$ и глубиной U_c (см. рис. 1). Пусть в яме существует не менее трех подзон размерного квантования, которые мы пронумеруем в порядке возрастания энергии как 1, 2 и 3. Предполагается, что квантовые ямы равномерно легированы.

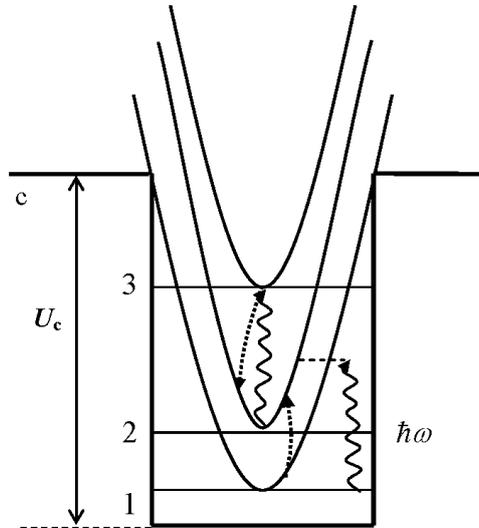


Рис. 1. Схема оптических и оже-переходов при фотонной лавине в легированных квантовых ямах, вертикальные волнистые линии со стрелками обозначают оптические переходы, пунктирные линии со стрелками – оже-переходы $31 \rightarrow 22$

В отсутствие оптической накачки электроны с концентрацией n_0 заполняют состояния вблизи дна нижней подзоны 1 до квазиуровня Ферми E_F , тогда как подзоны 2 и 3 практически не заселены. Предполагается, что энергетические зазоры между подзонами $\hbar\omega_{ij}$ велики по сравнению с E_F и T . Считаем также, что $\omega_{32} > \omega_{21}$, причем $\hbar(\omega_{32} - \omega_{21}) > E_F, T$, а частота падающего света ω попадает в резонанс с переходом между второй и третьей подзонами: $\omega \approx \omega_{32}$. При малых интенсивностях света j идут лишь очень слабые фотопереходы в области далекого коротковолнового крыла полосы поглощения между подзонами 1 и 2. Эти переходы являются непрямыми в двумерном \mathbf{k} -пространстве. Они происходят в состояния, далекие от дна подзоны 2, с передачей большого поперечного импульса, например, за счет участвующих в элементарном акте фононов. При увеличении j те немногие электроны, которые оказались в подзоне 2, быстро (за времена $\sim 10^{-13}$ с) попадают на дно этой подзоны, после чего могут либо опуститься еще ниже и вернуться в подзону 1, либо поглотить фотон $\hbar\omega$ и оказаться в подзоне 3. Сила осциллятора для резонансных разрешенных переходов $2 \rightarrow 3$ очень велика, т.к. определяется геометрическими размерами квантовой ямы. Из подзоны 3 электроны могут «свалиться» в подзоны 2 и 1. В то же время, как показано в [3], большой эффективностью обладает и процесс оже-типа $31 \rightarrow 22$: столкновение электрона в подзоне 3 с электроном в подзоне 1 приводит к тому, что они оба попадают в подзону 2. Каждый из

этих электронов может таким же образом привести к появлению двух электронов в подзоне 2 и т.д. При больших интенсивностях света скорость прихода в подзону 2 электронов, попадающих в нее благодаря этому механизму, превышает скорость их ухода в подзону 1 за счет межподзонной релаксации. В этом случае и происходит лавинообразное увеличение заселенности.

1.2. Гетероструктура типа II с квантовыми ямами

На рис. 2 показана схема переходов для энергетической структуры системы с квантовыми ямами типа II. Структура состоит из компонент *A* и *B*. Область с шириной $2a$, занимаемая компонентой *A* гетероструктуры (область *A*), является прямоугольной ямой для электронов с глубиной U_c и прямоугольным барьером для дырок высотой U_v . Считаем, что глубина ямы U_c достаточно велика ($\sim 1,5-2$ эВ). Вне указанной области находится компонента *B* (область *B*). В яме для электронов имеются три подзоны размерного квантования, пронумерованные в порядке возрастания энергии как 1, 2 и 3.

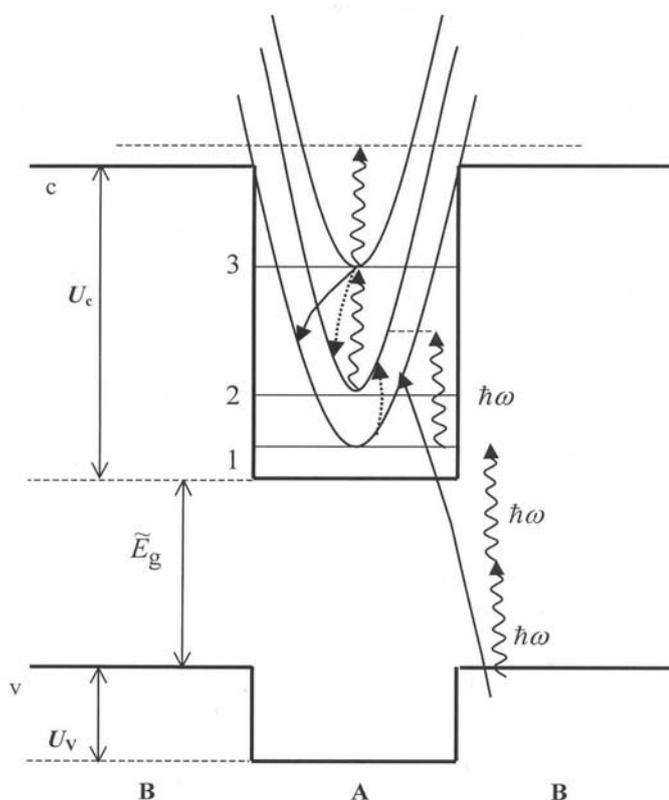


Рис. 2. Схема переходов в квантовой яме типа II при фотонной лавине, пунктирные линии со стрелками – оже-переходы $31 \rightarrow 22$, сплошные линии со стрелками – оже-переходы $3 \rightarrow 11v$, другие обозначения – как на рис. 1

В отличие от ситуации, рассмотренной в п. 1.1, предполагается, что в равновесных условиях электронные состояния в квантовой яме не заселены. Считается, что энергетические зазоры между подзонами $\hbar\omega_{ij}$ велики по сравнению с температурой T . Предполагается также, что $\omega_{32} > \omega_{21}$, а частота падающего света $\omega \approx \omega_{32}$. При малых интенсивностях света j появляется лишь небольшое число неравновесных электронов в подзоне 1 за счет однофотонных либо двухфотонных не прямых (в \mathbf{r} -пространстве) переходов из состояний валентной зоны v в области *B*. Поскольку частота света больше частоты прямых переходов ω_{21} между подзонами 1 и 2, между этими подзонами могут идти лишь слабые не прямые (в \mathbf{k}_\perp -пространстве) переходы. Быстрые резонансные фотопереходы могли бы идти между подзонами 2 и 3, но эти подзоны при малых j остаются

практически пустыми. Картина резко меняется при высоких значениях j . За счет двухступенчатого каскада слабых переходов $\nu \rightarrow 1$ и $1 \rightarrow 2$ некоторое число электронов все же оказывается в состояниях с $\mathbf{k}_\perp \neq 0$ подзоны 2. Далее развивается фотонная лавина, как это описывается выше. Подчеркнем, что ядром описанного механизма запуска фотонной лавины в случае легированной квантовой ямы является сильное поглощение света на переходах между возбужденными состояниями системы в сочетании с оже-переходом $31 \rightarrow 22$, ведущим к размножению электронов в подзоне 2.

В рассматриваемой модели для гетероструктуры типа II появляется и играет существенную роль еще один процесс оже-типа: электрон из подзоны 3 взаимодействует с электроном из валентной зоны области B и они оба попадают в подзону 1. При этом увеличивается общее число электронов в квантовой яме, что приводит к снижению пороговой интенсивности света j_{th} . Этот процесс оже-типа в сочетании с процессом $31 \rightarrow 22$, обеспечивающим, в частности, переход электронов из подзоны 1 в подзону 3, образует ядро еще одного лавинного механизма.

1.3. Гетероструктура типа I с квантовыми ямами

Зонная схема гетероструктуры типа I и существенные для фотонной лавины переходы в этом случае представлены на рис. 3. Зонная структура состоит из компонент A и B , где область A с шириной $2a$ является прямоугольной ямой как для электронов (с глубиной U_c), так и для дырок (с глубиной U_v). Предполагается, что глубина ямы U_c достаточно велика, а количество подуровней и энергетические расстояния между ними аналогичны случаю для гетероструктуры типа II. Считается, что $U_v < U_c$ и в яме для дырок имеется лишь одна подзона размерного квантования ν_1 . Развитие процесса фотонной лавины происходит так же, как описано в п. 1.2.

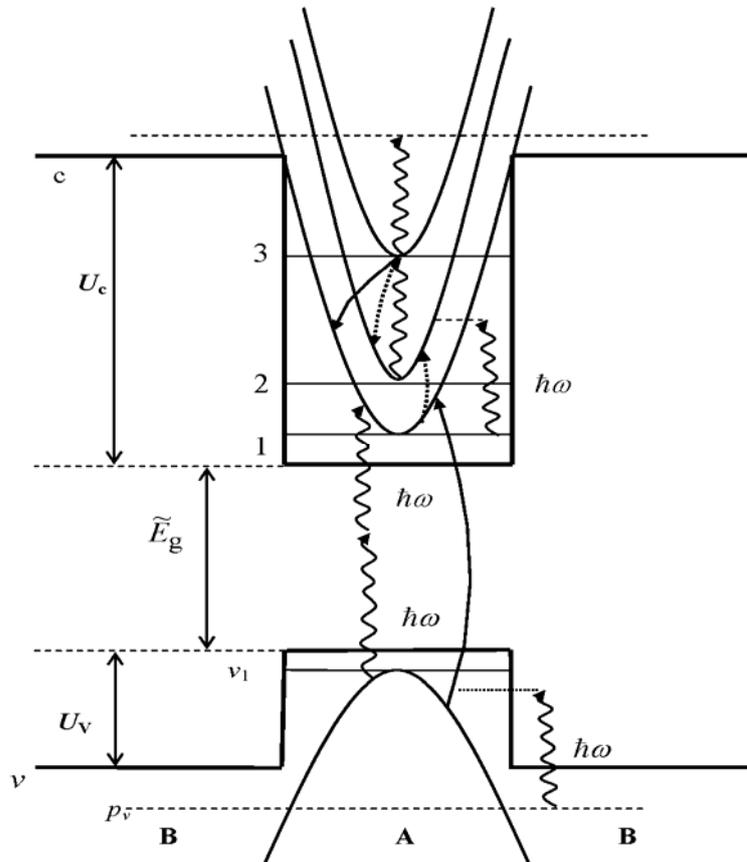


Рис. 3. Схема переходов в системе с квантовыми ямами типа I, обозначения такие же, как на рис. 2

В данной модели существенную роль играет процесс оже-типа $3 \rightarrow 11\nu_1$: электрон из подзоны 3 переходит в подзону 1, передавая высвободившуюся энергию на рождение пары, состоящей из электрона в подзоне 1 и дырки в подзоне ν_1 . Этот процесс отличается от случая для гетероструктуры типа II тем, что теперь обе начальные частицы находятся в квантовой яме (область A). В рассмотрение включаются также фотопереходы дырок из подзоны ν_1 (область A) в состояния непрерывного спектра валентной зоны (область B).

2. Система балансных уравнений

Математическое моделирование фотонной лавины в квантовых ямах проводится с использованием уравнений баланса для заселенностей электронных состояний, которые составляются для каждой описанной выше модели с учетом указанных переходов и процессов релаксации и рекомбинации носителей заряда. При составлении уравнений баланса следует иметь в виду, что при больших интенсивностях накачки j концентрации электронов в подзонах 2 и 3 становятся сопоставимыми, и следует учитывать фотоиндуцированные переходы как с поглощением, так и с испусканием фотона. Сказанное не относится к переходам $\nu \rightarrow 1$ (гетероструктура типа II), $p_\nu \rightarrow \nu_1$, $\nu_1 \rightarrow 1$ (гетероструктура типа I), $1 \rightarrow 2$ и $3 \rightarrow c$, так как конечные состояния для этих переходов практически не заполнены из-за отдаленности от минимумов соответствующих зон и подзон в квантовой яме. Обладая относительно большой кинетической энергией, электроны или дырки в этих конечных состояниях быстро теряют ее за счет внутризонной или внутриподзонной релаксации и покидают область резонанса.

Система уравнений для концентраций неравновесных дырок непрерывного спектра валентной зоны p_ν , дырок p_1 в квантовой яме валентной зоны, электронов $n_{1,2,3}$ в трех подзонах квантовой ямы зоны проводимости и электронов в непрерывном спектре зоны c n_c для модели (3) имеет вид

$$\begin{aligned} \dot{n}_c &= \sigma_{3c} j n_3 - W_{c3} n_c - d_c n_c (p_0 + p), \\ \dot{n}_3 &= W_{c3} n_c - \sigma_{3c} j n_3 - (W_{31} + W_{32} + \tau_{\text{aug}}^{-1}) n_3 + \sigma_{23} j (n_2 - n_3) - W_{31,22} (n_1, n_3), \\ \dot{n}_2 &= -W_{21} n_2 - \sigma_{23} j (n_2 - n_3) + W_{32} n_3 + 2W_{31,22} (n_1, n_3) + \sigma_{12} j n_1, \\ \dot{n}_1 &= \sigma_{12} j n_1 + W_{21} n_2 + (W_{31} + 2\tau_{\text{aug}}^{-1}) n_3 - W_{31,22} (n_1, n_3) + \\ &\quad + \sigma_{p_1} j^2 (1 - f_{p_1}) (1 - f_{c_1}) - d_1 n_1 (p_{10} + p_1), \\ \dot{p}_1 &= d_1 n_1 (p_{10} + p_1) + \sigma_{p_1} j^2 (1 - f_{p_1}) (1 - f_{n_1}) + \tau_{\text{aug}}^{-1} n_3 + \sigma_{p_\nu p_1} p_\nu j - W_{p_1 p_\nu} p_1, \\ \dot{p}_\nu &= d_c n_c (p_0 + p_\nu) - \sigma_{p_\nu p_1} p_\nu j + W_{p_1 p_\nu} p_1 \end{aligned} \quad (1)$$

с начальными условиями: $n_1, n_2, n_3, n_c = 0, p_1 = p_{01}, p_\nu = p_0$ при $t = 0$. В формулах (1) $W_{c3}, W_{p_1 p_\nu}$ – скорости захвата электронов и дырок из непрерывного спектра зоны проводимости в подзону 3 квантовой ямы зоны проводимости и подзону p_1 квантовой ямы валентной зоны; $W_{ij} (i > j)$ и σ_{ij} – скорость релаксации и сечение оптического перехода между i -й и j -й подзонами; $W_{31,22}$ – вероятность оже-перехода $31 \rightarrow 22$, а величина τ_{aug}^{-1} описывает скорость оже-перехода $3 \rightarrow 11\nu_1$.

В уравнениях (1) фигурируют «двумерные» концентрации p_ν, p_0, n_c частиц непрерывного зонного спектра, которые отличаются от обычных концентраций $\tilde{p}_\nu, \tilde{p}_0, \tilde{n}_c$ множителем $n_w = 5 \cdot 10^5 \text{ см}^{-1}$ (количество ям на единицу длины вдоль оси роста наноструктуры): $\tilde{n}_c = n_c n_w$ и т.п. Члены $-d_c n_c (p_0 + p_\nu)$ и $-d_1 n_1 (p_{01} + p_\nu)$ описывают обычную бимолекулярную рекомбинацию электронов в зоне проводимости и в подзоне 1 с дыр-

ками в валентной зоне и в подзоне p_1 области A . Величины p_0, p_{10} обозначает равновесную концентрацию дырок в непрерывном спектре валентной зоны и в подзоне p_1 . Равновесную концентрацию электронов в зоне проводимости считаем равной нулю.

При высоких интенсивностях света состояния вблизи дна нижней подзоны размерного квантования квантовой ямы для электронов оказываются заполненными, а состояния вблизи потолка подзоны в яме для дырок опустошаются (т.е. возникает высокая концентрация дырок p). Это влияет на скорости межзонных переходов и переходов оже-типа. Данные эффекты не являются критическими для рассматриваемой задачи. Тем не менее, их желательно учесть хотя бы в самом грубом приближении. С этой целью реальные распределения неравновесных электронов и дырок аппроксимируются фермиевскими функциями распределения, соответствующими мгновенным концентрациям $n_1(t)$ и $p_1(t)$ электронов и дырок. Так, в формулах (1) фигурируют функции распределения f_{n_1} и f_{p_1} электронов и дырок в подзоне n_1 и p_1 , соответствующие энергиям $\varepsilon_c(\mathbf{k}_\perp)$ и $\varepsilon_v(\mathbf{k}_\perp)$ в точке \mathbf{k}_\perp двухфотонного резонанса между подзонами:

$$f_m = \left\{ \exp\left[\left(\varepsilon_m(\mathbf{k}_\perp) - \mu_m(m)\right)/k_B T\right] + 1 \right\}^{-1}, \quad (2)$$

где k_B – постоянная Больцмана, $m = n_1, p_1$.

Для вычисления химических потенциалов $\mu_{n_1}, \mu_{p_1}, (\mu_{p_1} > 0)$ для электронов и дырок, фигурирующих в формуле (2) для функций распределения, воспользуемся трансцендентными уравнениями

$$\frac{n_1(t)}{N_{n_1}(T)} - \Phi(\mu_{n_1}^*) = 0, \quad \frac{p_1(t)}{N_{p_1}(T)} - \Phi(\mu_{p_1}^*) = 0, \quad (3)$$

где $\Phi_{1/2}(x) = \int_0^\infty \frac{dz}{1 + \exp(z - x)}$, $\mu_m^* = \mu_m^*/k_B T$, $N_{n_1} = \frac{m_c k_B T}{\pi \hbar^2}$ и $N_{p_1} = \frac{m_v k_B T}{\pi \hbar^2}$ – эффективные плотности состояний, соответственно, в подзоне 1 квантовой ямы зоны проводимости и подзоне в яме валентной зоны, m_i – эффективная масса частицы в зоне i .

Аналогичным образом для двух других моделей составляются системы балансных уравнений, которые здесь не приводятся. Однако следует отметить, что система уравнений баланса для легированных квантовых ям является наиболее простой из рассматриваемых моделей, что позволяет провести ее качественный анализ (см. [4, 5]). В случае гетероструктуры типа I требуется получить самосогласованное решение системы нелинейных дифференциальных уравнений (1) и трансцендентных уравнений (3).

3. Оценка параметров системы

Получение решения системы уравнений (1) связано с необходимостью оценки значений входящих в нее параметров. Для этого требуется: (а) рассчитать (или оценить по имеющимся экспериментальным данным) вероятности фотопереходов между различными электронными состояниями квантовой ямы, (б) рассчитать скорости «лавинно-образующих» процессов – процессов оже-типа в квантовой яме, (в) оценить скорости релаксационных процессов. К числу последних относятся излучательные и безызлучательные процессы, причем в большинстве актуальных случаев процессы второго типа играют доминирующую роль. В случае материалов с КЯ основную роль в процессах внутри- и межподзонной релаксации играет взаимодействие с полярными оптическими фононами. Анализ усложняется тем, что необходимо учитывать взаимодействие как с конфайнментными, так и с интерфейсными колебательными модами. Этим вопросам посвящена довольно обширная литература (см. например, [6–11]). Сечения захвата носителей из зонных состояний в подзону размерного квантования в квантовой яме удается оценить лишь очень грубым образом, равно как и сечения межзонной рекомбина-

ции неравновесных электрон-дырочных пар. Впрочем, значения этих величин не являются критическими для построения качественной картины эффекта фотонной лавины.

3.1. Вероятности двухфотонных переходов и переходов оже-типа

Для оценки вероятности прямых (в координатном пространстве и в пространстве поперечных импульсов) двухфотонных переходов между подзонами была использована модель с однозонными волновыми функциями (см. [12]) и учитывалась специфика правил отбора для межзонных и межподзонных оптических переходов. Вычисления проводились во втором порядке теории возмущений по взаимодействию электронной подсистемы со светом. Были получены аналитические выражения для вероятности двухфотонных переходов через интенсивность света, приведенную массу электрона и дырки, межзонный матричный элемент оператора импульса электрона и превышение энергии двух фотонов над порогом двухфотонных переходов.

Расчет скорости процессов оже-типа в квантовой яме оказывается технически весьма сложным, однако удастся получить достаточно надежные оценки скоростей этих процессов [5, 12–14].

Рассмотрим переходы оже-типа $3 \rightarrow 11\nu_1$: электрон с двумерным волновым вектором $\mathbf{k}_{3\perp}$, находящийся в подзоне 3 квантовой ямы зоны проводимости, переходит в подзону 1 в состоянии с волновым вектором $\mathbf{k}_{11\perp}$. Высвободившаяся энергия расходуется на рождение пары, состоящей из электрона в подзоне 1 с волновым вектором $\mathbf{k}_{12\perp}$ и дырки в подзоне ν_1 с волновым вектором $-\mathbf{k}_{\nu\perp}$. Переданный при этом импульс в плоскости, перпендикулярной оси роста наноструктуры, равен:

$$\mathbf{q}_{0\perp} = \mathbf{k}_{3\perp} - \mathbf{k}_{11\perp} = \mathbf{k}_{12\perp} - \mathbf{k}_{\nu\perp}.$$

Выражение для вероятности данного перехода запишем в виде

$$W(\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}) = \frac{2\pi}{\hbar} \sum_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}, \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}} f(\mathbf{k}_{3\perp}) \left| M_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(d)} + M_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(e)} \right|^2 \times \delta(E_{3\mathbf{k}_{3\perp}} + E_{\nu\mathbf{k}_{\nu\perp}} - E_{1\mathbf{k}_{11\perp}} - E_{1\mathbf{k}_{12\perp}}),$$

где $f(\mathbf{k}_{3\perp})$ – функция распределения электронов в подзоне 3 зоны проводимости, а $M_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(d)}$ и $M_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}; \mathbf{k}_{11\perp}, \mathbf{k}_{12\perp}}^{(e)}$ – прямой и обменный матричные элементы оператора межэлектронного взаимодействия. При написании формулы для вероятности предполагалось, что электронные состояния $\mathbf{k}_{11\perp}$, $\mathbf{k}_{12\perp}$ и дырочные состояния $-\mathbf{k}_{\nu\perp}$ не заняты. Это обусловлено, с одной стороны, большой величиной расстройки резонанса $\Delta'_\nu = (E_{03} - 2E_{01} - \tilde{E}_g - E_{0p_1})$, (E_{0i} – энергия экстремума i -й подзоны размерного квантования, \tilde{E}_g – см. на рис. 3) и короткими временами релаксации электронов в подзоне размерного квантования, а с другой – уходом дырок под действием излучения накачки в состояния непрерывного спектра валентной зоны).

Подставляя кулоновское взаимодействие в виде разложения в трехмерный ряд Фурье, после стандартных преобразований в результате получим следующее выражение для прямого матричного элемента:

$$M_{\mathbf{k}_{3\perp}, \mathbf{k}_{\nu\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{\nu\perp} + \mathbf{q}_{0\perp}}^{(d)}(q_{0\perp}) = \frac{2e^2}{\varepsilon_l S} \chi_{\parallel} \left[i\tilde{I}_1(q_{0\perp}) + \sqrt{2} \cos \theta_{\mathbf{q}_{0\perp}\pi_{\perp}} \tilde{I}_2(q_{0\perp}) \right],$$

где $\chi = \frac{\hbar\pi_{cv}}{mE_g}$, $\pi = \mathbf{p} + \frac{\hbar}{4mc^2} [\sigma \nabla V_L]$, σ – матрица Паули, S – площадь квантовой ямы,

V_L – периодический потенциал решетки, \mathbf{p} – оператор импульса, $\theta_{\mathbf{q}_{0\perp}\pi_{\perp}}$ – угол между векторами $\mathbf{q}_{0\perp}$ и π_{\perp} , а величины $\tilde{I}_i(q_{0\perp})$ представляют собой следующие выражения:

$$\begin{aligned}\tilde{I}_1(q_{0\perp}) &= \iint dz_1 dz_2 e^{-q_{0\perp}|z_1-z_2|} \text{sign}(z_1-z_2) \beta_3^*(z_1) \beta_1(z_1) \beta_{v1}^*(z_2) \beta_1(z_2), \\ \tilde{I}_2(q_{0\perp}) &= \iint dz_1 dz_2 e^{-q_{0\perp}|z_1-z_2|} \beta_3^*(z_1) \beta_1(z_1) \beta_{v1}^*(z_2) \beta_1(z_2),\end{aligned}$$

где $\beta_i(z)$ – огибающие волновые функции подзон размерного квантования.

Выражение для вероятности перехода запишем в виде

$$\begin{aligned}W_{3,11v} &= \frac{2^3 S^3}{(2\pi)^5 \hbar} \int d^2 \mathbf{k}_{v\perp} \int f(\mathbf{k}_{3\perp}) d^2 \mathbf{k}_{3\perp} \int d^2 \mathbf{q}_{0\perp} \left| M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{v\perp} + \mathbf{q}_{0\perp}}^{(d)} + \right. \\ &+ \left. M_{\mathbf{k}_{3\perp}, \mathbf{k}_{v\perp}, \mathbf{k}_{3\perp} - \mathbf{q}_{0\perp}, \mathbf{k}_{v\perp} + \mathbf{q}_{0\perp}}^{(e)} \right|^2 \delta \left[\Delta'_v + \frac{\hbar^2}{m_c} (\mathbf{k}_{3\perp} - \mathbf{k}_{v\perp}) \cdot \mathbf{q}_{0\perp} - \frac{\hbar^2 q_{0\perp}^2}{m_c} - \frac{\hbar^2 k_{v\perp}^2}{2m_r} \right].\end{aligned}\quad (4)$$

Для упрощения вычисления шестикратного интеграла в выражении для вероятности (4) будем считать, что все электроны в подзоне 3 находятся вблизи ее дна. Такое предположение обусловлено тем обстоятельством, что релаксация электронов внутри подзоны 3 происходит значительно быстрее других релаксационных процессов. Тогда в подынтегральной функции в правой части (4) можно положить $\mathbf{k}_{3\perp} = 0$. При этом

$$\int f(\mathbf{k}_{3\perp}) d^2 k_{3\perp} = (2\pi)^2 n_3.$$

Интегрирование по углам между $\mathbf{k}_{3\perp}$ и \mathbf{q}_0 снимается с помощью δ -функции. После некоторых вычислений получим:

$$\begin{aligned}\frac{1}{\tau_{\text{aug}}} &= \frac{W_{3,11v}}{S n_3} = \frac{4 e^4 m_c \chi_{\parallel}^2}{\pi^2 \hbar^3 a^2 \varepsilon_t^2} \int_0^{k_{v\perp}^{(m)}} dk_{v\perp}^2 \int_{\mathcal{G}_-}^{\mathcal{G}_+} \frac{dq_0^2 \left| \tilde{M}_{\mathbf{q}_{0\perp}}^{(d)} + \tilde{M}_{\mathbf{q}_{0\perp}}^{(e)} \right|^2}{\sqrt{(q_{0\perp}^2 - \mathcal{G}_-)(\mathcal{G}_+ - q_{0\perp}^2)}}, \\ \mathcal{G}_{\pm} &= \frac{1}{2} (k_{v\perp} \pm \rho)^2, \quad \rho = \sqrt{k_{v\perp}^2 + \frac{4m_c a^2}{\hbar^2} \left(\Delta'_v - \frac{\hbar^2 k_{v\perp}^2}{2m_r} \right)}, \\ k_{v\perp}^{(m)} &= \frac{a^2}{\hbar^2} \frac{4m_v m_c}{2m_c + m_v} \Delta'_v.\end{aligned}\quad (5)$$

В формулах (5) фигурируют безразмерные волновые векторы, измеряемые в единицах a^{-1} , где a – постоянная решетки.

3.2. Оптимизация параметров

Оптимизация параметров зонной структуры для уменьшения пороговой интенсивности света, необходимой для запуска фотонной лавины, проводилась в пределах, обусловленных (а) соотношениями между энергетическими параметрами, допускаемыми выбранной моделью эффекта, и (б) значениями параметров, встречающихся в реально существующих структурах. В результате были выбраны следующие значения параметров, фигурирующих в правых частях уравнений (1): $a = 3 \cdot 10^{-7}$ см, $p_{01} = 0$, $p_0 = 3 \cdot 10^{10}$ см⁻², $W_{31} = 0,02$ пс⁻¹, $W_{p_1 p_v} = W_{c3} = 0,01$ пс⁻¹, $W_{32} = 0,07$ пс⁻¹, $W_{21} = 0,01$ пс⁻¹, $\tau_{\text{aug}} = 0,06$ пс, $\delta = 0,05$ эВ, $\alpha = 1,2$, $m_v = 0,5m$, $m_c = 0,04m$, $\sigma_{23} = 2$ см²/пс⁻¹МВт, $\sigma_{12} = 0,003$ см²/пс⁻¹МВт, $\sigma_{p_1 p_i} = \sigma_{3c} = 0,025$ см²МВт, $U_v = 0,3$ эВ, $E_g = 1$ эВ, $T = 300$ К, $\sigma_{p_v n_i} = 4,7 \times 10^4$ см²МВт⁻²пс⁻¹, $d_1 = 0,01$ см²с⁻¹, $d_c = 0,003$ см²с⁻¹, $U_c = 1,74$ эВ.

Результаты численных расчетов

Результаты численного решения системы нелинейных уравнений (1) даны на рис. 4 и 5. На рис. 4 приведены зависимости квазиравновесных заселенностей

$n_{1,2,3,c}$ и $p_{1,v}$ от интенсивности накачки. Видно, что имеется пороговое значение интенсивности $j = j_{th}$, вблизи которого происходит резкое увеличение заселенностей подзон в квантовой яме для электронов, подзоны в яме для дырок, непрерывного спектра в зоне проводимости и валентной зоне. В данном случае $j_{th} \approx 0,159 \text{ МВт/см}^2$.

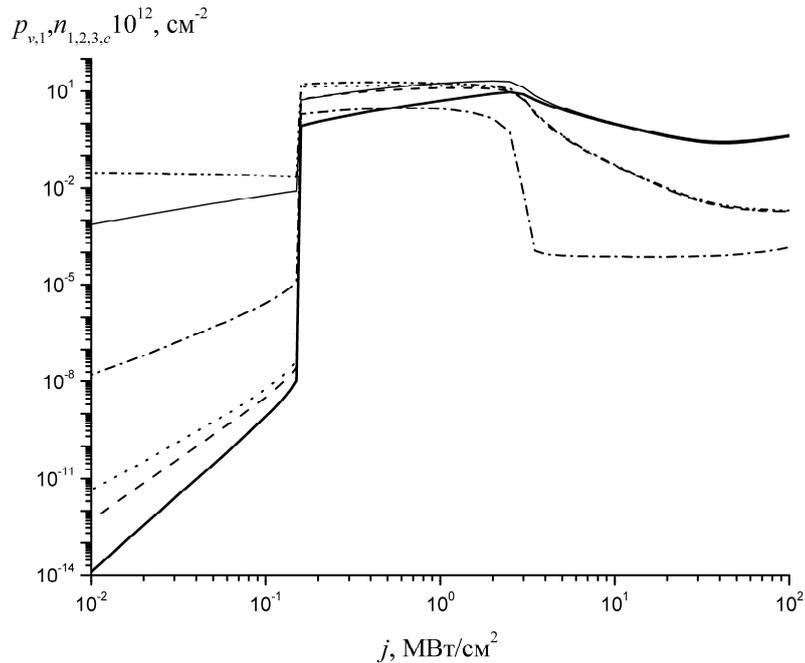


Рис. 4. Зависимость равновесных заселенностей от интенсивности света, где жирная сплошная линия – n_c , штриховая – n_3 , пунктирная – n_2 , штрих-пунктирная – n_1 , штрих два пунктира – $p_{1,v}$, сплошная – p_v

На рис. 5 изображены зависимости времени установления квазиравновесного распределения электронов τ_{eq} от интенсивности накачки j . Видно, что вблизи $j = j_{th}$ величины τ_{eq} резко возрастают. Указанные эффекты являются следствием возникновения фотонной лавины (см. Введение).

Схожие с приведенными на рис. 4, 5 результаты получены также для систем с легированными квантовыми ямами [5] и с глубокими квантовыми ямами типа II [12, 13]. Таким образом, анализ кинетики всех трех моделей показывает характерные для эффекта фотонной лавины особенности. Вместе с тем, каждая из рассмотренных моделей имеет уникальные характеристики, что определяет целесообразность практического использования модели в конкретной ситуации.

Сопоставим оптическое переключение и ап-конверсию в квантовых ямах различных типов. В случае легированных квантовых ям при начальной (двумерной) концентрации свободных электронов в нижней подзоне размерного квантования $\sim 10^{12} \text{ см}^{-2}$ времена переключения в актуальном диапазоне интенсивностей света ($\sim 1 \text{ МВт/см}^2$) могут составлять от единиц до десятков пикосекунд, а энергии переключения весьма малы ($\sim 10\text{--}100 \text{ фДж/мкм}^2$). Весьма низкими являются и пороговые интенсивности света ($j_{th} \sim 10\text{--}100 \text{ кВт/см}^2$). Гетероструктуры с подходящими зонными параметрами достаточно распространены. Существенным недостатком в случае данной схемы является невысокая степень ап-конверсии: фактически можно получить люминесценцию (и лазерную генерацию) на частоте, менее чем в 2 раза превосходящей частоту накачки.

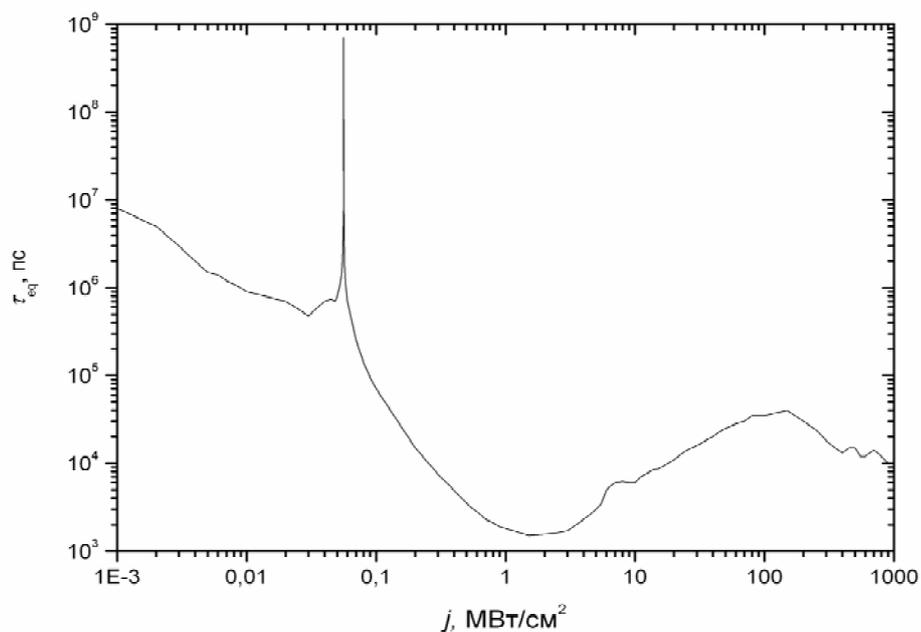


Рис. 5. Зависимость времени установления равновесных заселенностей τ_{eq} от интенсивности света

Этот недостаток преодолевается в моделях, использующих двухфотонную накачку нижней подзоны размерного квантования в яме для электрона в гетероструктурах типов I и II. Так, в структуре с квантовыми ямами типа II можно получить люминесценцию на межзонных переходах на частоте, почти в 5 раз превышающей частоту накачки. При этом пороговые интенсивности света остаются примерно такими же, как в случае легированных квантовых ям. Однако из-за того, что в такой модели требуется время для создания достаточно высокой концентрации носителей в нижней подзоне размерного квантования, времена (а вместе с ними и энергии) переключения оказываются примерно на два порядка большими, чем в случае эффекта фотонной лавины в материале с легированными квантовыми ямами. Примером гетероструктуры типа II с глубокими КЯ могут служить системы $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ (область A)/ $\text{AlAs}_{0.56}\text{Sb}_{0.54}$ (см. [15]) либо $\text{In}_{0.3}\text{Ga}_{0.7}\text{As}$ (область A)/ AlAs (область B) (см. [16]).

В случае гетероструктуры типа I с глубокими квантовыми ямами степень апконверсии может превышать 5, а время (~ 1 нс) и энергия переключения (~ 1 пДж/см²) принимают значения, промежуточные между теми, которые можно реализовать в моделях легированных квантовых ям и квантовых ям типа II. В принципе такой вариант мог бы считаться оптимальным, однако в настоящее время известна лишь одна гетеросистема CdSe/SrS [17], зонные параметры которой соответствуют требованиям данной модели.

Заключение

В работе исследована кинетика фотоиндуцированных переходов в структурах с глубокими квантовыми ямами с различными типами электронной зонной структуры. Для описания динамики системы составлены уравнения баланса для концентрации частиц в подзонах размерного квантования квантовой ямы. Решение системы нелинейных дифференциальных уравнений для каждой из рассматриваемых моделей показывает наличие характерных для эффекта фотонной лавины особенностей. На основе полученных результатов проведен сравнительный анализ моделей.

Литература

1. Chivian J.S., Case W.E., Eden D.D. The photon avalanche: a new phenomenon in Pr^{3+} -based infrared quantum counters // *Appl. Phys. Lett.* – 1979. – V. 35. – № 2. – P. 124–136.
2. Joubert M.-F. Photon avalanche upconversion in rare earth laser materials // *Optical materials.* – 1999. – V. 11. – P. 181–202.
3. Перлин Е.Ю. Фотонная лавина в легированной квантовой яме // *Оптика и спектроскопия.* – 2001. – Т. 5. – В. 5. – С. 777–783.
4. Perlin E.Yu. Photon Avalanche Effect in Doped Quantum Wells // *Journ. of Luminescence.* – 2001. – V. 94–95. – P. 249–253.
5. Перлин Е.Ю., Левицкий Р.С. Фотонная лавина в легированных квантовых ямах: упр-конверсия и эффект переключения // *Оптический журнал.* – 2006. – Т. 73. – В. 1. – С. 3–11.
6. Degani M.H. and Hipolito O. Electron-phonon interaction effects in a quasi-two-dimensional electron gas in the $\text{GaAs-Ga}_{1-x}\text{Al}_x\text{As}$ heterostructure // *Phys. Rev. B.* – 1987. – V. 35. – № 14. – P. 7717–7722.
7. Ridley B.K. Electron scattering by confined LO polar phonons in a quantum well // *Phys. Rev. B.* – 1989. – V. 39. – No 8. – P. 5282–5289.
8. Ridley B.K. Electron-hybridon interaction in a quantum well // *Phys. Rev. B.* – 1993. – V. 47. – No 8. – P. 4592–4594.
9. Nash K.J. Electron-phonon interactions and lattice dynamics of optic phonons in semiconductor heterostructures // *Phys. Rev. B.* – 1992. – V. 46. – № 12. – P. 7723–7729.
10. Trallero-Giner C., Comas F. and Garsia-Moliner F. Polar optical modes and electron-phonon interaction in semiconductor nanostructures // *Phys. Rev. B.* – 1994. – V. 50. – № 3. – P. 1755–1764.
11. Gondar J.L., Comas F. and Castro F. Scattering rates in a semiconductor heterostructure: the effects of intersubband transitions // *Physica B.* – 2000. – V. 292. – P. 354–356.
12. Перлин Е.Ю., Иванов А.В., Левицкий Р.С. Каскадно-лавиная генерация электрон-дырочных пар в квантовых ямах типа II // *ЖЭТФ.* – 2003. – Т. 123. – В. 3. – С. 612–624.
13. Перлин Е.Ю., Иванов А.В., Левицкий Р.С. Каскадно-лавиная ап-конверсия и генерация неравновесных электрон-дырочных пар в гетероструктурах типа II с глубокими квантовыми ямами // *Оптический журнал.* – 2006. – Т. 73. – В. 1. – С. 12–21.
14. Левицкий Р.С., Иванов А.В., Перлин Е.Ю. Эффект фотонной лавины в гетероструктурах типа I с глубокими квантовыми ямами // *Оптический журнал.* – 2006. – Т. 73. – В. 2. – С. 3–8.
15. Neogi A., Yoshida H., Mozume T. et al. Absorption saturation of near-infrared intersubband transition in lattice-matched InGaAs/AlAsSb quantum wells // *Physica E.* – 2000. – V. 7. – P. 183–187.
16. Garcia C.P., De Nardis A., Pellegrini V. et al. 1.26 μm intersubband transitions in $\text{In}_{0.3}\text{Ga}_{0.7}\text{As/AlAs}$ quantum wells // *Appl. Phys. Lett.* – 2000. – V. 77. – № 23. – P. 3767–3769.
17. Engelmann R., Ferguson J. and Solanki R. Quantum-well activated phosphors: A new concept for electroluminescent displays // *Appl. Phys. Lett.* – 1997. – V. 70. – № 4. – P. 411–413.

ИССЛЕДОВАНИЕ СТРУКТУРЫ И СВОЙСТВ ОПТИЧЕСКИХ УФ-ОТВЕРЖДАЕМЫХ АКРИЛАТНЫХ ПОЛИМЕРОВ, НАПОЛНЕННЫХ НАНОЧАСТИЦАМИ ZNO

С.А. Семьина

Научный руководитель – Ю.Э. Бурункова

Исследована наномодификация УФ-отверждаемых акрилатных композитов при введении структурирующей нанодобавки ZnO и ее влияние на оптические свойства и влагопоглощение. Обнаружен эффект аномального снижения влагоемкости более чем в 5 раз при введении 10% наночастиц в композит при сохранении его прозрачности. Рельеф поверхности и внутренняя структура полученных нанокомпозитов исследованы методами атомно-силовой микроскопии.

Введение

Применение полимеров во многих оптических устройствах обусловлено тем, что они дешевы, технологичны, а их эксплуатационные свойства можно изменять в зависимости от области применения. Однако вопросы оптической однородности и стабильности свойств полимерных материалов по сравнению с традиционными материалами требуют дальнейшего изучения.

Взаимодействию воды с полимерами и изучению в этих системах процессов массопереноса уделяется большое внимание, так как вода является наиболее распространенной контактной средой, с которой приходится сталкиваться при хранении, производстве и эксплуатации. Влагопоглощение оптического полимера приводит к необратимому ухудшению его оптических характеристик и является основным фактором, ограничивающим их применение. Знание закономерностей процессов диффузии и сорбции воды необходимо для оценки возможности использования полимерных материалов в условиях длительного нахождения в атмосфере паров воды, разработки научно-обоснованных методов оценки работоспособности оптических устройств на основе полимеров. Кроме того, существует аномальность и неаддитивность сорбционных свойств таких систем, например, отрицательная концентрационная зависимость коэффициентов диффузии для одних полимеров и постоянная для других.

Свойства полимеров зависят от химического строения и размеров макромолекул и от взаимного их расположения (структуры). Согласно двумерной модели фибриллы со складчатыми доменами, дискретная структура аморфных полимеров описывается следующим образом: участки макромолекулярных цепей доменов уложены параллельно друг другу, домены соединены проходными цепями, плотность которых значительно ниже, чем внутри домена. Весь полимер представляет собой набор фибрилл, которые взаимодействуют между собой. Наличие пространства между макромолекулами приводит к тому, что часть объема тела остается незанятой его массой, так называемый свободный объем [1, 2].

Водяные пары могут адсорбироваться на внешней и внутренней поверхности полимеров или растворяться в их микропорах и более мелких пустотах, возникающих между макромолекулами. Взаимодействие паров органических веществ или воды с полимерами зависит от их химической природы. Неполярные полимеры лучше сорбируют пары неполярных веществ, полярные полимеры – пары полярных веществ и воды [3]. При этом происходят существенные изменения структуры полимеров, которые набухают и даже растворяются. Плотность упаковки макромолекул всегда оказывает влияние на проницаемость – при более рыхлой упаковке наблюдается большая сорбция. Поэтому любые факторы, изменяющие характер упаковки (введение низкомолекулярных веществ), способствуют изменению сорбции полимеров.

Исследованию вопросов диффузии и сорбции воды индивидуальными полимерами посвящено достаточно большое число работ [4–6]. Однако проблемы, возникающие при

введении низкомолекулярных добавок, которые осложняют анализ сорбции и изменение оптических свойств полимеров, являются вопросом, требующим детального освещения.

1. Постановка задачи

Основная задача работы состоит в исследовании влагопоглощения как частного вопроса проницаемости полимерных композиций в зависимости от свободного объема и влияния, которое оказывает наличие паров воды в системе на оптические характеристики полимерного композита.

Уменьшение влагопоглощения чрезвычайно важно для разработки полимерных интегрально-оптических устройств, поскольку повышенное влагопоглощение приводит к ухудшению и нестабильности практически всех характеристик: потери света в результате светорассеяния и изменения показателя преломления, нестабильность характеристик электрооптических полимеров, существенное понижение лучевой прочности и т.д.

В работе исследован способ решения проблемы влагопоглощения полимерной композиции путем введения в полимер наночастиц ZnO, заполняющих свободный объем и структурирующих матрицу.

2. Используемые вещества

Для создания полимерной матрицы в работе использовались вещества, представленные в табл. 1. Для УФ-отверждения использовали инициатор 2,2-диметокси-2-фенилацетонфенон (Aldrich 19,611-8) в концентрации 0,2 вес. %. В качестве структурирующей добавки были использованы наночастицы ZnO размером 20 нм.

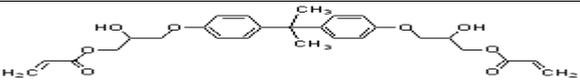
Название	Сокращение	Химическая формула
2 карбокси этилакрилат	2Carb	$\text{H}_2\text{C}=\text{CH}-\overset{\text{O}}{\parallel}{\text{C}}-\text{OCH}_2\text{CH}_2-\overset{\text{O}}{\parallel}{\text{C}}-\text{OH}$
бисфенол А глицеролат	BisA	
3 метилол пропан этоксилат диакрилат	TMP	$(\text{H}_2\text{C}=\text{CHCO}_2\text{CH}_2\text{CH}_2\text{OCH}_3)_3\text{C}(\text{C}_2\text{H}_5)\text{CH}_2\text{OCH}_2\text{CH}_2\text{OCH}_3$

Таблица 1. Используемые мономеры

3. Методы исследований

В работе влагопоглощение изучается гравиметрическим методом. Метод заключается в следующем: образцы выдерживались при постоянной температуре (50°C) и влажности 95–98% в течение суток в термокамере. Затем производили взвешивание с точностью 0,001г. Влагопоглощение (α ,%) образцов рассчитывали по изменению массы (ΔM), отнесенному к исходной (M):

$$\alpha = \left(\frac{\Delta M}{M}\right)100 \%$$

Использовалось 3–8 образцов для каждого вида полимера.

Рассеяние света на полимерах измеряется в соответствии с рекомендациями европейского стандарта ASTM D1003. По методике этого стандарта светорассеяние образцов определяется как относительная доля излучения, которое при прохождении через обра-

зец отклоняется на угол более 2,5 угл. град. от направления падающего на образец хорошо сколлимированного зондирующего пучка. При этом учитывается рассеяние как от поверхностей, ограничивающих испытываемый образец, так и его внутренние части. Измерения проводятся с использованием фотометрического шара.

Спектры пропускания пленок были измерены на спектрофотометре Perkin-Elmer 555.

4. Результаты и их обсуждение

4.1. Эксперимент

Значения влагопоглощения свободных полимеров и их композиций представлены в табл. 2.

состав	$\alpha, \%$ экспериментальное	$\alpha, \%$ расчетное
TMP	7,4	-
2 Carb	56,9	-
BisA	1,37	-
2Carb/BisA (70/30)	23,10	26,52
TMP/2Carb (50/50)	35,31	32,15

Таблица 2. Влагопоглощение полимеров и их бинарных композиций

Как видно из табл. 2, величина сорбции для бинарных композиций не является аддитивной. Это можно объяснить тем, что при введении второго полимера происходит переформирование исходной структуры, неаддитивно изменяется свободный объем.

Нами предложен способ решения проблемы уменьшения влагопоглощения путем введения в полимер наполнителя, изменяющего свободный объем – наночастиц (ZnO).

Создан нанокомпозит на основе полимера и структурирующей добавки порошкообразного ZnO (до 10 вес. %). Совместимость ZnO с системой BisA/2Carb наблюдается до 14 вес. %. При введении в мономерную композицию Tmp/2Carb более 10 вес. % частиц ZnO при полимеризации образуются непрозрачные пленки, происходит расслоение системы, она становится гетерогенной.

Исследуемые образцы нанокомпозита, как видно из рис. 1, представляют собой прозрачные пленки в УФ и видимой области спектра.

Также следует заметить, что спектры пропускания пленок, полученных на основе BisA/2Carb и Tmp/2Carb, различаются. Это обусловлено различной упаковкой молекул в указанных составах и разной совместимостью мономерных компонентов с ZnO. Пленки состава BisA/2Carb более прозрачны в видимой области спектра при введении наночастиц ZnO, а пропускание полимерной композиции Tmp/2Carb заметно уменьшается с увеличением концентрации ZnO более 10 вес. %. По-видимому, структурирующая добавка ZnO менее совместима с составом Tmp/2Carb.

При введении ZnO в составы мономерной композиции величины влагопоглощения изменяются немонотонно (рис. 2). Подобный характер изменения свойств полимерной композиции позволяет предположить, что происходит переформирование структуры (укладки макромолекул) в результате введения структурирующей нанодобавки ZnO, аналогично тому, что наблюдалось в работе [7].

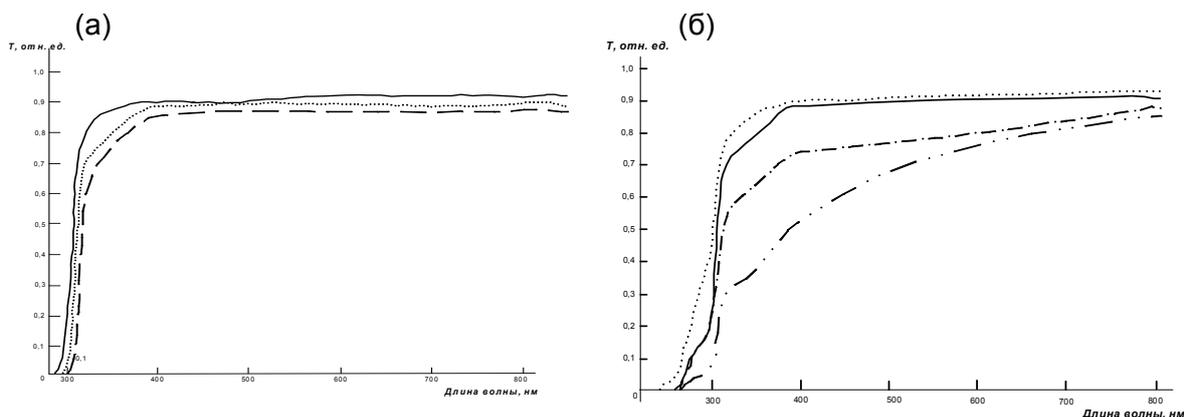


Рис. 1. Спектры пропускания пленок составов BisA/2Carb (а) и Tmp/2Carb (б) при введении ZnO: — 0 вес. %;4 вес. %; _._._7 вес.%; _._._. 10 вес. % ZnO

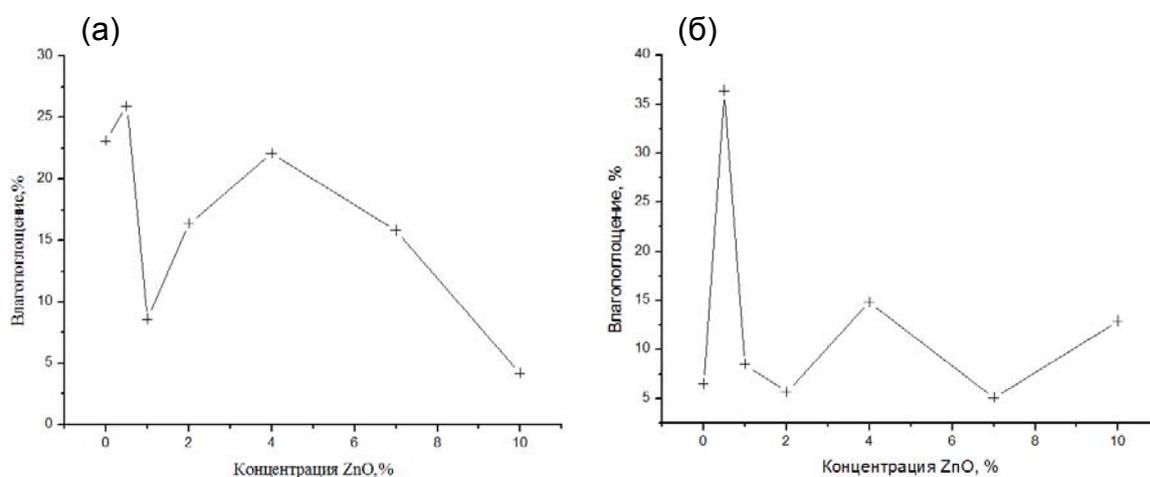


Рис. 2. Зависимость внутренней сорбции полимерной композиции 2Carb/Bis A (а) и 2Carb/Tmp (б) от концентрации наночастиц ZnO

Из рис. 2 видно, что при концентрациях до 4 вес. % концентрации ZnO недостаточно для равномерного распределения их по всему объему полимера для обоих составов, поэтому структура композита имеет рыхлый неоднородный характер, что отражается на изменениях величины сорбции (наличие экстремумов).

При введении в полимер свыше 4 вес. % наночастиц (рис. 2) величины сорбции для обоих составов изменяются по-разному.

В композиции BisA/2Carb при концентрации частиц ZnO, начиная с 4 % и до 10 %, идет равномерное уменьшение сорбции, т.е. происходит такое изменение укладки макромолекул, при котором свободный объем системы монотонно уменьшается. По сравнению с влагопоглощением ненаполненного композита, при введении в композит BisA/2Carb 10 вес. % ZnO сорбция воды уменьшается практически в пять раз. Для составов Tmp/2Carb в данной области концентраций величина сорбции практически не изменяется, а по сравнению с влагопоглощением ненаполненного композита сорбция воды даже возрастает.

Поскольку данные полимерные составы часто используются при создании микроэлементов в интегрально-оптических схемах, было целесообразным исследовать их величину светорассеяния. На рис. 3 представлены зависимости величины светорассеяния полимеров от концентрации введенных наночастиц ZnO до и после испытаний на влагопоглощение.

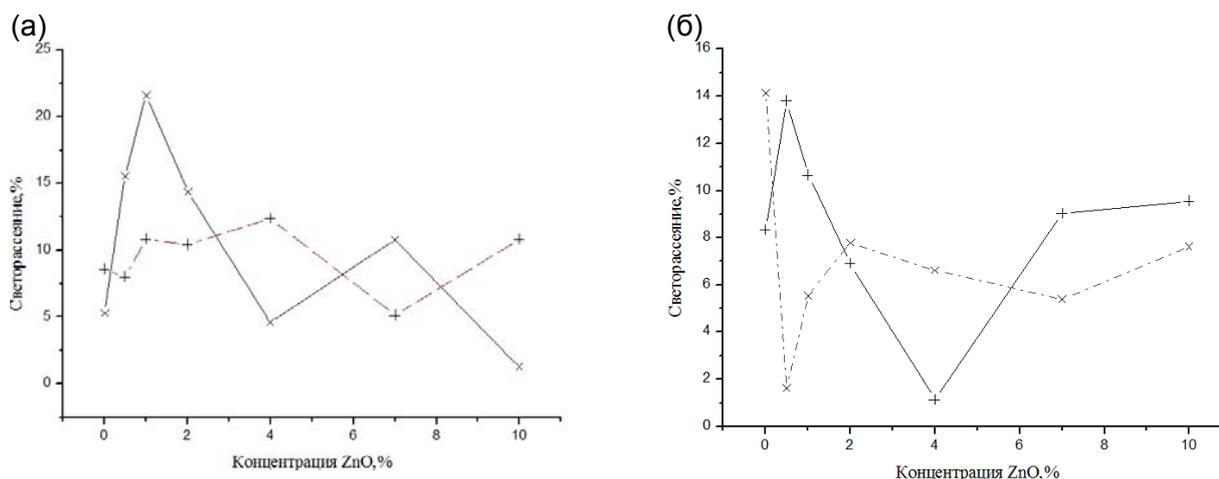


Рис. 3. Зависимость светорассеяния полимерной композиции 2Carb/BisA (а) и 2Carb/Tmp (б) от концентрации наночастиц ZnO до(—) и после (----) влагопоглощения

В области концентрации наночастиц до 4 вес. % для обоих составов величина светорассеяния изменяется немонотонно. При концентрации 1 вес. % величина светорассеяния максимальна, а затем резко уменьшается. При содержания наночастиц в полимере более 4 вес. %, в зависимости от состава, изменение светорассеяния пленок имеет различный характер. Для состава BisA/2Carb наблюдается незначительное изменение величины светорассеяния, и при наибольшей концентрации ZnO – 10 вес. % – светорассеяние составляет примерно 2,5 %, что практически не отличается от светорассеяния ненаполненного композита. Для пленок состава Tmp/2Carb после концентраций 4 вес. % ZnO величины светорассеяния резко увеличиваются и далее практически не изменяются, достигая 11 % при 10 вес. % ZnO, тогда как для чистого композита рассеяние составляет 8 % .

После испытаний пленок на влагопоглощение характер изменения рассеяния света в зависимости от содержания наномодификатора для состава BisA/2Carb резко изменяется: плавно возрастает с увеличением содержания ZnO до 4 вес. %, а затем практически не меняется. Для состава Tmp/2Carb сохраняется резко немонотонное изменение свойств до 2 вес. % добавки и далее практически не меняется. Такое изменение светорассеяния после испытаний пленок на влагопоглощение может объясняться тем, что при повышенной температуре в водной атмосфере активно идут релаксационные процессы, и происходит переупаковка макромолекул в более равновесное состояние.

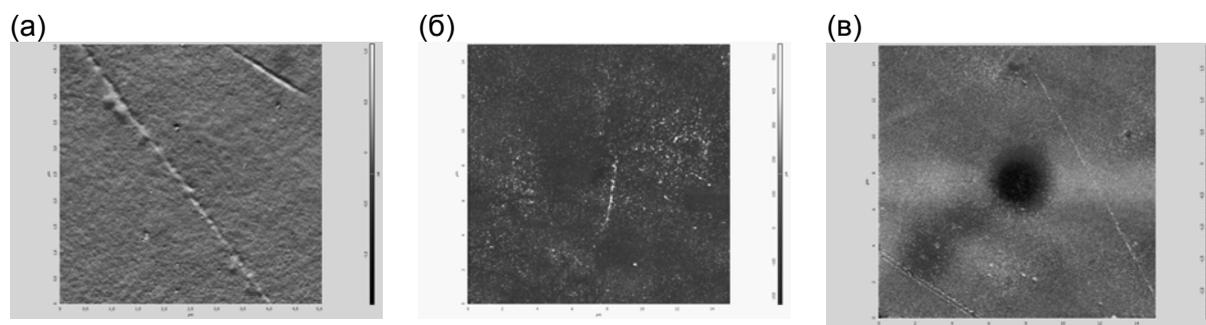


Рис. 4. Топография поверхности пленок состава BisA/2Carb (30/70) при введении 0 вес. % (а), 4 вес. %, 10 вес. % наполнитель ZnO

Наблюдаемая корреляция изменения влагопоглощения пленок с уровнем светорассеяния подтверждает наши предположения о структурирующем воздействии нанодобавки ZnO на взаимное расположение макромолекул полимеров в обоих составах. Надо заметить, что для составов BisA/2Carb добавка ZnO активнее изменяет свойства композиции, чем для составов с Tmp/2Carb.

Исследование рельефа поверхности и внутренней структуры полимера при введении наночастиц проведено методами атомно-силовой микроскопии (рис.4). Для фотографий использовались образцы пленок размерами 15×15 мкм. Как видно, уже при введении 4 вес. % наночастиц ZnO происходят существенные изменения структуры композиции по сравнению с исходной для чистого полимера. Отчетливо наблюдается формирование отдельных полимерных областей, структурированных ZnO. При 10 вес. % ZnO мы видим уже иную укладку макроцепей, равномерную по всему материалу. Подобные изменения полимерной матрицы наблюдаются и для состава Tmp/2Carb.

Заключение

В работе исследовались два широко используемых для изготовления микроэлементов в интегрально-оптических схемах полимерных состава: BisA/2Carb и Tmp/2Carb с введением и без наночастиц ZnO. Изучались изменения таких характеристик, как сорбция, пропускание и светорассеяние в зависимости от концентрации вводимого в исследуемые композиции наночастиц ZnO. Пленки состава BisA/2Carb более прозрачны в видимой области спектра при достаточно значительных концентрациях наночастиц ZnO, а пропускание полимерной композиции Tmp/2Carb быстро уменьшается с увеличением концентрации ZnO свыше 4 вес. %. Введение ZnO в полимер немонотонно изменяет величины влагопоглощения и светорассеяния нанокompозита. Подобный характер изменения свойств полимерной композиции позволяет предположить, что происходит перестроение структуры (укладки макромолекул) в результате возможного взаимодействия структурирующей добавки ZnO с активными группами полимера.

Характер зависимости светорассеяния и влагопоглощения от концентрации наночастиц для BisA/2Carb позволяет заключить, что средство этого полимерного состава к ZnO лучше, чем композит состава Tmp/2Carb.

Получено уменьшение влагопоглощения композита по сравнению с чистым полимером в 5 раз для состава BisA/2Carb при сохранении уровня светорассеяния.

Литература

1. Денисюк И.Ю., Фокина М.И., Бурункова Ю.Э. Полимеры в интегральной и волоконной оптике – физика, технология и применение. – СПб: СПбГУИТМО, 2008. – 89 с.
2. Гуль В.Е. Структура и прочность полимеров. – М.: Химия, 1971. – 344 с.
3. Тагер А.А. Физико-химия полимеров. – М.: Химия, 1978. – 543 с.
4. Кротов А.С. Диагностика процессов сорбции и диффузии влаги в полимерных композиционных материалах: Автореф. дис. ... канд. хим. наук. – Барнаул, 2002. – 20 с.
5. Прокофьева Т.А., Давыдова Е.В., Майорова Н.В. Влияние сорбции низкомолекулярных веществ на структуру сшитых полиэфиров. // Высокомолекулярные соединения. – 1980. – Т. XXII А. – №1. – С. 174.
6. Мамаева И.А., Давыдова Е.В., Сорбция воды полиэфирмеламиновыми пленками сетчатого строения // Высокомолекулярные соединения. – 1985. – Т. XXVII А. – №4. – С. 844.
7. Гинзбург Б.М., Табаров С.Х., Туйчиев Ш., Шепелевский А.А. Влияние добавок фуллерена C60 на структуру и механические свойства тонких пленок из органического стекла // Письма в ЖТФ. – 2007. – Т. 33. – Вып. 23. – С. 43.

КОНЦЕНТРАЦИОННОЕ ТУШЕНИЕ ЛЮМИНЕСЦЕНЦИИ ЕВРОПИЯ В ОРГАНО-НЕОРГАНИЧЕСКОМ КОМПОЗИТЕ

О.А. Миноженко

Научный руководитель – В.А. Асеев

Золь-гель методом синтезированы органо-неорганические композиты, в которых органические комплексы европия ковалентно связаны с кремнеземной матрицей, и изучены их люминесцентные свойства.

Введение

Большинство переходных металлов и лантаноидов образуют комплексы с органическими соединениями, содержащими гетероатомы с неподелёнными электронными парами, такими как азот и кислород. При использовании лигандов, способных образовывать хелатные структуры, например 2,2'-бипиридила, 1,10-фенантролина, 1,3-дикарбонильных соединений, устойчивость этих комплексов существенно повышается [1]. Многообразие электронных переходов в таких комплексных соединениях, обусловленное доступностью d и f электронных подуровней переходных металлов и лантаноидов, в сочетании с сопряженными π -электронными системами вышеуказанных лигандов обуславливает их интересные фотофизические, фотохимические и электрохимические свойства [2–6]. Обычно при исследовании этих свойств в качестве модельных соединений используют комплексы европия, обладающие интенсивной и устойчивой люминесценцией. При этом оптимальными являются комплексы с четырьмя лигандами, во внутренней координационной сфере которых отсутствуют молекулы воды, приводящей к тушению возбуждённых состояний системы. С точки зрения растворимости комплексов предпочтительнее работать с системами со смешанными лигандами, например трис(2-теноилтрифторацетонато)-1,10-фенантролин европием (рис. 1) и аналогичными структурами.

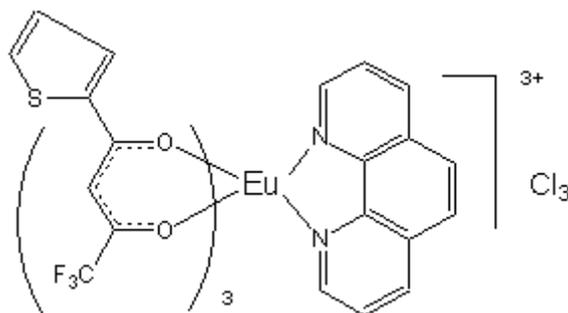


Рис. 1. Трис (2-теноилтрифторацетонато) -1,10-фенантролин европий

Практическое использование подобных соединений связано с необходимостью их внедрения в матрицу. В качестве матрицы могут применяться разнообразные органические полимеры – полиметакрилаты, полиуретаны, полиамиды, полиимиды, полибензимидазолы и т.д. [7–12], а также неорганические оксиды ряда элементов, синтезируемые золь-гель методом [13, 14].

Наиболее простым является подход, при котором комплексы лантаноидов внедряются в термически и химически стабильную матрицу SiO_2 . Устойчивый интерес к таким материалам обусловлен перспективами их использования в качестве твердотельных лазеров, электролюминесцентных устройств, антибликовых покрытий, оптических волокон и нелинейно-оптических материалов [15–17].

Образование органо-неорганического гибрида в системе комплекс РЗЭ – кремнезем происходит как в результате межмолекулярных взаимодействий компонентов системы [9, 10], так и их химического связывания [11]. Естественно, что во втором случае

образуется более стабильная и однородная по составу структура, в которой органический компонент системы равномерно распределяется в объеме материала [18, 19]. Отсутствие равномерного распределения органического компонента в неорганической матрице приводит к локальным увеличениям его концентрации и образованию кластеров лантаноидов, что фатальным образом сказывается на люминесцентных свойствах синтезированных материалов. При гомогенном распределении органического компонента в системе вероятность кластерообразования, а, следовательно, и концентрационного тушения люминесценции существенно понижается [20].

Практической реализации химического связывания комплексов лантаноидов с кремнеземной матрицей посвящен ряд публикаций [21–23]. Наиболее простым является введение в комплекс РЗЭ производного 1,10-фенантролина, содержащего триалкоксисилильные группы. В ряде работ для получения такого производного использовали реакцию между 5-амино-1,10-фенантролином и триэтоксисилилпропилизотиоцианатом [21, 22]. Однако ввиду низкой основности атома азота аминогруппы в гетероциклическом амине протекание подобного процесса представляется маловероятным. При этом если замещение первого атома водорода аминогруппы ещё можно осуществить, то из-за малой активности алифатического изотиоцианата замещение атома водорода в образовавшейся мочеvine достаточно сомнительно. К тому же продолжительное нагревание реагентов приводит к существенному осмолению реакционной смеси. В связи с этим был предложен принципиально новый подход к синтезу триалкоксисилильного производного 1,10-фенантролина (рис. 2), в котором исходными реагентами являлись активный гетероциклический изотиоцианат и алифатический амин [24] (рис. 2).

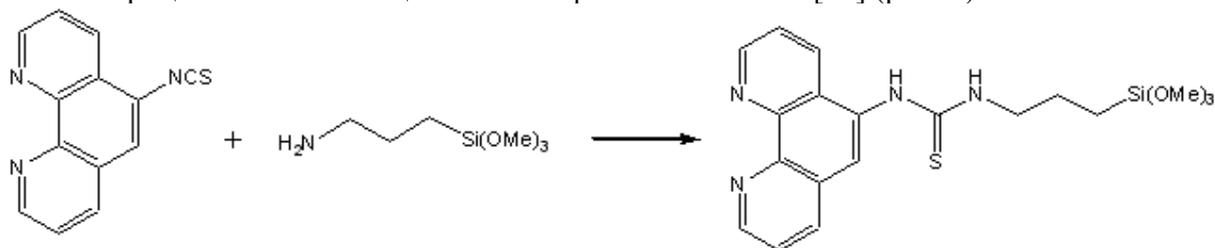


Рис. 2. Синтез триалкоксисилильного производного 1,10-фенантролина

Реакция между этими соединениями протекает быстро и количественно уже при комнатной температуре, что важно для получения исходных комплексов (рис. 3) в чистом виде.

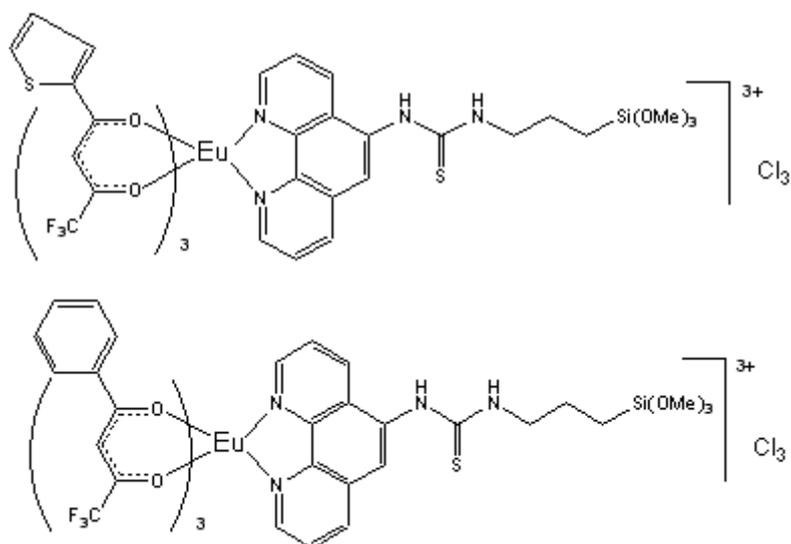


Рис. 3. Исходные комплексы

На основе комплексов (рис. 3) были сформированы органо-неорганические гибриды, компоненты которых химически связаны друг с другом. Были исследованы люминесцентные свойства европия в полученных органо-неорганических композитах.

Методика исследований

Исследуемые образцы отличались друг от друга концентрацией европия в композите, концентрационный ряд представлен в таблице.

№ образца	Концентрация европия
1	0.06 вес. %
2	0.13 вес. %
3	0.31 вес. %
4	0.62 вес. %
5	0.08 вес. %

Таблица. Концентрационный ряд

Спектры люминесценции были получены на спектрофлуориметре Perkin Elmer LB-50B, возбуждение на длине волны 532 нм.

Для исследования концентрационного тушения люминесценции перехода ${}^5D_0 - {}^5F_2$ ионов европия в органо-неорганическом композите использовалась установка (рис. 4). Излучение лазера 1 возбуждает люминесценцию в образце 2, которая регистрируется приемником 4, и на осциллографе 5 наблюдается кривая распада люминесценции.

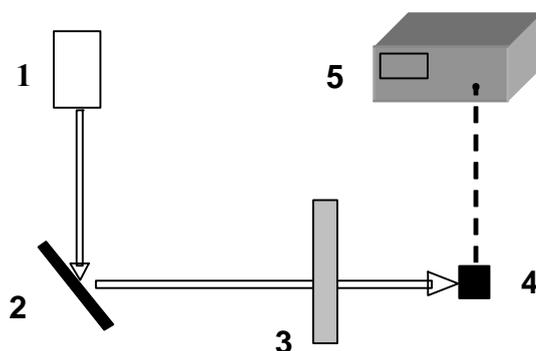


Рис. 4. Схема установки для исследования концентрационного тушения люминесценции: 1 – импульсный лазер, 2 – образец, 3 – спектральный фильтр, 4 – приемник, 5 – осциллограф

Результаты и их обсуждение

На рис.5 показан полученный спектр люминесценции образца с процентным содержанием европия 0,8 вес. %.

Из графика видно, что максимум полосы излучения образца находится на длине волны ~610 нм, что соответствует переходу ${}^5D_0 - {}^7F_2$. Полуширина спектральной линии составляет ~90 нм.

Для исследования концентрационного тушения люминесценции европия в органо-неорганическом композите по методике, описанной выше, были измерены времена жизни для образцов различной концентрации. Полученные результаты представлены на рис. 6.

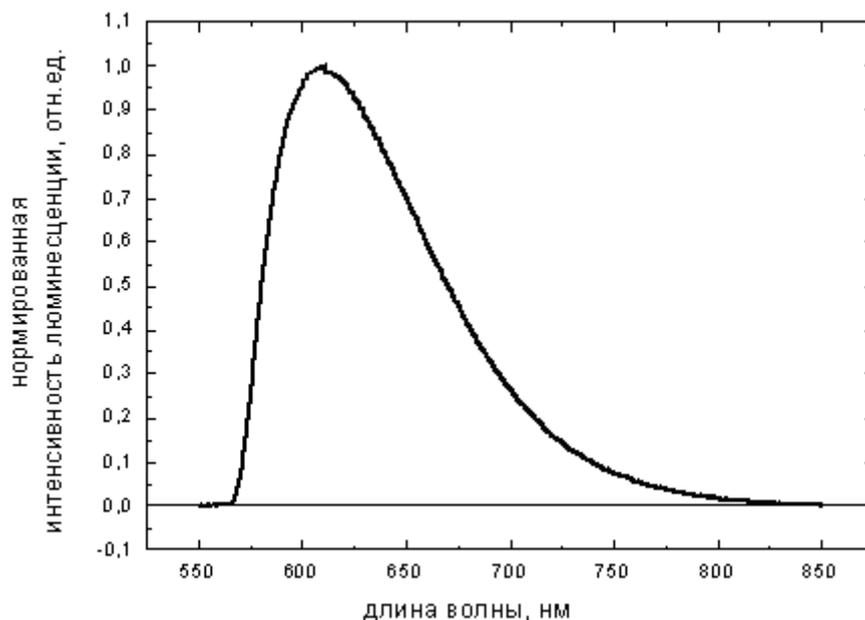


Рис. 5. Спектр люминесценции образца №5

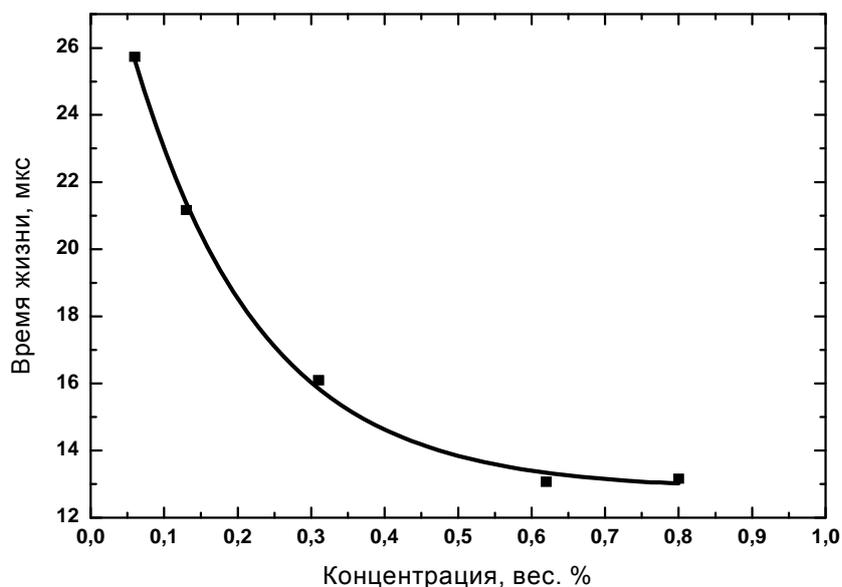


Рис. 6. Концентрационная зависимость времени жизни

Можно сделать вывод о том, что концентрационная зависимость времени затухания люминесценции имеет нелинейный характер, а увеличение концентрации в ~ 10 раз приводит к уменьшению времени жизни люминесценции в ~ 2 раза.

Заключение

В данной работе были проведены экспериментальные исследования люминесцентных свойств европия в органо-неорганическом композите. Были получены следующие результаты: спектры люминесценции у полученных образцов представляют собой полосу с полушириной ~ 90 нм и максимумом на длине волны 610 нм, что соответствует переходу 5D_0 - 7F_2 , а концентрационная зависимость времени затухания люминесценции европия в органо-неорганическом композите имеет нелинейный характер.

Литература

1. Кермак В., Мак-Кейл Д. Гетероциклические соединения. / Под ред. Р. Эльдерфилда. – М.: Мир, 1965. – Т.7. – С. 287.
2. Lo K.K.-W., Chan-Ming Ng D., Chung C.-K. // *Organometallics*. – 2001. – V. 20. – P. 4999–5002.
3. Tyson D.S., Castellano F.N. // *Inorg. Chem.* – 1999. – V. 38. – P. 4382–4386.
4. Lo K.K.-W., Chung C.-K., Zhu N. // *Chem. Eur. J.* – 2003. – V. 9. – P. 475–480.
5. Yang J., Gordon K., Zidon Y., Shapira Y. // *J. Appl. Phys.* – 2003. – V. 94. – № 10. – P. 6391–6395.
6. Земский В.И., Вересов А.В., Ершов А.Ю. // *Опт. и Спектр.* – 1996. – Т. 81. – № 2. – С. 251–257.
7. Younathan J.N., McClanahan S.F., Meyer T.J. // *Macromolecules*. – 1989. – V. 22. – P. 1048–1052.
8. Rasmussen S.C., Thompson D.W., Singh V.S., Petersen J.D. // *Inorg. Chem.* – 1996. – V. 35. – P. 3449–3450.
9. V.C. Costa, W.L. Vasconcelos, K.L. Bray. // *J. Sol-Gel Sci. And Techn.* – 1998. – № 13. – P.605–609.
10. Ikeda S., Senna M. // *J. Mater. Chem.* – 2004. – № 14. – P. 1071–1074.
11. R.B. Pode. // *Phys. Stat. Sol. (a)*. – 1998. – V.170. – P. 137–144.
12. R.A. Ferreira, L.D. Carlos, R.R. Gonsales et al. // *Chem. Mater.* – 2001. – № 13. – P. 2991–2998.
13. Brinker C.J., Scherer G.W. / *Sol-Gel Science: the Physics and Chemistry of Sol-Gel Processing*. - San Diego: Acad. Press, 1990. 908 p.
14. *Sol-Gel Science And Technology.V. I. Processing Characterisation and Applications.* / Ed. Sakka S. N.Y.: Kluver Academic Publishes. 2005. 680 p.
15. Shiu F.-M., Chen M.-H., Tang R.-F., Jeng Y.-J., Chang M.-Y., Perng J.-H. // *J. Non-Cryst. Solids*. – 1997. – V. 209. – № 1. – P. 61–68.
16. Clark A., Terpigov V., Medrano F., Cervantes M., Soto D. // *Optical Materials*. – 1999. – V. 13. – P. 355–360.
17. Bekiari V., Lianos P. // *J. Luminescence*. – 2003. – V. 101. – P. 135–140.
18. Shavaleev N.M., Moorcraft L.P., Pope S.J.A. et al. // *Chem. Eur. J.* – 2003. – V. 9. – P. 5283–5285.
19. Химич Н.Н., Сёмов М.П., Чепик Л.Ф. // *ДАН.* – 2004. – Т. 394. – № 5. – P. 636–638.
20. Левшин В.Л. Фотолюминесценция жидких и твердых веществ. – М.: Изд. Техн.-теор. литературы, 1951. – 456 с.
21. H.R. Li, L.S. Fu, J. Lin, H.J. Zhang. // *Thin Solid Films*. – 2002. – № 416. – P. 197–200.
22. K. Binnemans, P. Lenaerts, K. Driesen, C. Goller- Warland. // *J. Mater. Chem.* – 2004. – № 14. P. 191–195.
23. H. Bauer, J. Blank, D.L. Ross. // *JACS*. – 1964. – V. 68. – № 23. – P. 5125–5131.
24. Химич Н.Н., Обрезков Н.П., Коптелова Л.А. // *ЖОрХ.* – 2006. – Т. 42. – Вып. 4. – С. 574–576.

НАНОИМПРИНТ – ФОРМИРОВАНИЕ НАНО- И МИКРО-ЭЛЕМЕНТОВ ФОТОНИКИ КОНТАКТНЫМ КОПИРОВАНИЕМ

Н.Н. Арефьева

Научный руководитель – д.ф.-м.н., профессор И.Ю. Денисюк

Исследованы процессы получения полимерных нано- и микро- оптических элементов методом контактного копирования с использованием гибкого штампа. На примере копирования линзы Френеля определена точность передачи формы при копировании. На примере копирования эшелеты 1800 лин./мм определена возможность копирования наноструктур.

Введение

В последнее время широко исследуются методы литографии, обеспечивающие возможность массового изготовления элементов, имеющих нанометровые размеры. С учетом ограниченности возможностей классической фотолитографии, определяющихся дифракционным пределом, на первый план выходят контактные методы литографии, обеспечивающие в перспективе возможность получения элементов размером вплоть до 1–1,5 нм [1].

В данной работе нами был реализован метод «soft lithography», являющийся разновидностью группы методов наноимпринтинга для формирования нано- и микро-элементов фотоники: решетки нанопризм (300 нм), дифракционные решетки, волноводные структуры, френелевские линзы. Метод основан на процессе двукратного копирования оригинала с получением промежуточной «зеркальной» копии на гибком штампе и второй копии на УФ-акрилате [2].

Изготовление гибкого штампа

Реализация метода начинается с изготовления гибкого штампа с имеющегося мастер-штампа (оригинала), выполненного на металле или фоторезисте. В качестве полимера для изготовления штампа использовался «Силастик Т-4» [3]. Основа «Силастик Т-4» представляет собой вязкую текучую прозрачную композицию на основе силиконового каучука, способную переходить в резиноподобное состояние после перемешивания с отвердителем. Отверждение происходит при комнатной температуре в течение 24 часов. Формы могут эксплуатироваться длительное время при температурах до +250°C. Итак, заливаем оригинал данной композицией. Проведенные эксперименты показали необходимость проведения обезгаживания силасстика, залитого на матрицу. При отсутствии обезгаживания в вакууме в силастике остаются пузыри воздуха, а также, вероятно, и растворенный воздух. В результате мелкие канавки и острые углы не прорабатываются. Напротив, при обезгаживании в вакууме 10^{-3} мм. рт. ст. в течение 30–40 мин. жидкий силастик кипит, что свидетельствует о выходе воздуха. После прекращения вакуумирования штамп вулканизируется на воздухе в течение 18–24 часов при комнатной температуре и отделяется от образца. Отделение происходит достаточно просто, так как силастик не имеет адгезии к подложкам и является гибким и прочным.

УФ-полимеризуемые композиции

В работе в качестве основного оптического материала используются УФ-полимеризуемые композиции. Композиция обычно состоит из олигомера, мономера и фотоинициатора. Преимуществами УФ-полимеризуемых мономерных композиций являются быстрое отверждение, обеспечивающее высокую производительность, отсутствие растворителей и выделяющихся летучих органических соединений.

Использованные компоненты УФ-отверждаемой композиции:

1. феноксиэтилакрилат (мономер);
2. бисфенол А глицеролат диакрилат (бифункциональный мономер, поперечно-сшивающий агент);
3. диметоксифенилацетофенон (инициатор).

Важной для оптических применений особенностью УФ-отверждаемых композиций является возможность варьирования показателя преломления в широких пределах при изменении соотношения компонентов. На рис. 1 показана возможность изменения показателя преломления УФ-бинарной композиции в диапазоне 1,49–1,54 при изменении соотношения компонентов [4].

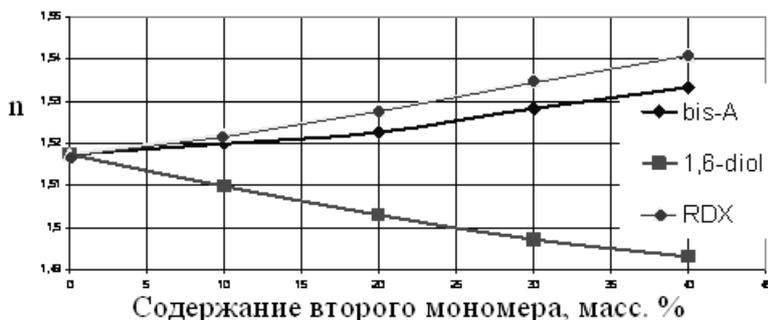


Рис. 1. Изменение показателя преломления бинарной композиции PEA – второй компонент в зависимости от их соотношения

Обозначения: PEA – феноксиэтилакрилат; 1,6-diol – 1,6-гександиол диакрилат; bis – Bisphenol A glycerolate; RDX – торговое название бромированного эпоксиакрилата фирмы UCB, США.

Соотношение компонентов бинарной композиции было выбрано для получения показателя преломления 1,54 и высокой твердости в отвержденном состоянии. Известно, что УФ-отверждаемые акрилаты ингибируются кислородом воздуха, и на воздухе реакция полимеризации не проходит [5]. Обычно для УФ-полимеризации без использования вакуума или аргона процесс проводят в закрытом объеме. Здесь штамп мог бы являться одной из поверхностей этого объема. Однако оказалось, что силастик прозрачен для кислорода, а его емкость по кислороду столь значительна, что УФ-отверждение акрилата, контактирующего со штампом, не начинается даже через 30 мин. после начала экспонирования. Только вакуумирование или напуск аргона позволяет начаться процессу фотополимеризации.

Штамп из силастика оптически прозрачен на длине волны более 280 нм, поэтому возможна полимеризация УФ-излучением при экспонировании через штамп. Реакцию полимеризации обеспечивает фотоинициатор, входящий в состав акриловой композиции, который активизируется под действием энергии фотонов и, расщепляясь, вступает в химическую реакцию с олигомерами и мономерами, сшивая их в длинные трехмерные цепочки полимера.

Метод наноимпринта

Рассмотрим реализацию метода наноимпринта на примере линзы Френеля рис. 2. Для экспериментальной реализации метода была изготовлена установка, фотография которой приведена на рис. 3.

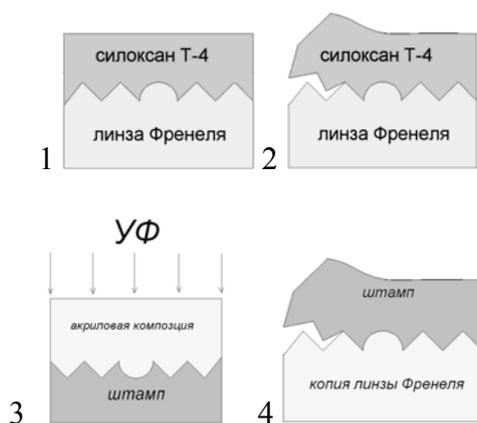


Рис. 2. 1 – изготовление гибкого штампа на основе силистика Т-4, 2 – Снятие штампа после полимеризации, 3 – Фотоотверждение на длине волны 365 нм, 4 – Снятие готовой линзы. Штамп готов к повторному использованию



Рис. 3. Лабораторная установка. 1 – УФ-осветитель, 2 – элемент прижима гибкого штампа и экспонирования, 3 – съемный вакуумный колпак

Установка представляет собой откачной пост, обеспечивающий откачку до форвакуума и проведение процесса в среде аргона. Установка позволяет реализовать все необходимые условия для реализации процесса «soft lithography» [6,7].

С использованием данного метода было выполнено копирование френелевской линзы и эшелеты 1800 лин/мм.

Для эксперимента в качестве образца была использована линза Френеля, изготовленная методом литья под давлением поликарбоната по стандартной технологии. Суть эксперимента заключается в создании копии данной линзы с помощью метода наноимпринта. Результаты эксперимента приведены на рис. 4.

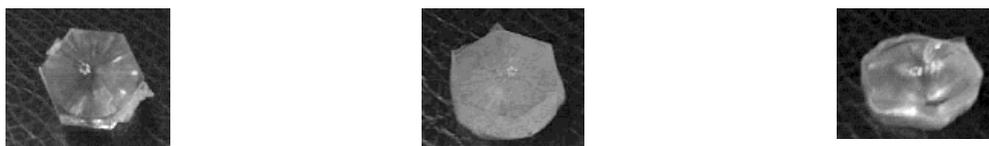


Рис. 4. Фотографии линзы-оригинала (ПММА), гибкого штампа (силоксан Т-4) и копии (УФ-акрилат)

Для контроля формы поверхности гибкий штамп был разрезан и разрез сфотографирован. Как видно из таблицы, сравнение полученных углов структуры с расчетными показало их идентичность с точностью 5–10 минут. Следовательно, данный метод позволяет копировать микрооптику.

N	α_N теор	α_N практ									
1	1°20'	1°20'	14	37°40'	37°35'	27	60°20'	60°15'	40	72°40'	72°40'
2	4°	4°	15	39°50'	39°50'	28	61°10'	61°10'	41	73°20'	73°20'
3	7°30'	7°35'	16	42°	42°	29	62°45'	62°40'	42	74°	74°
4	10°50'	10°55'	17	44°10'	44°10'	30	63°50'	63°50'	43	74°40'	74°40'
5	13°30'	13°30'	18	46°10'	46°10'	31	64°55'	64°55'	44	75°17'	75°20'
6	16°30'	16°30'	19	48°10'	48°10'	32	66°10'	66°20'	45	75°55'	75°50'
7	19°30'	19°30'	20	50°	50°	33	67°	67°	46	76°33'	76°35'
8	22°30'	22°30'	21	52°10'	52°10'	34	67°50'	67°50'	47	77°12'	77°12'
9	25°30'	25°25'	22	54°	54°	35	68°45'	68°45'	48	78°05'	78°10'
10	28°10'	28°10'	23	55°20'	55°20'	36	69°40'	69°40'	49	79°10'	79°10'
11	30°30'	30°35'	24	56°30'	56°30'	37	70°40'	70°45'	50	80°05'	80°05'
12	32°50'	32°40'	25	57°50'	57°50'	38	71°20'	71°15'			
13	35°10'	35°10'	26	59°10'	59°10'	39	72°05'	72°05'			

Таблица. Сравнение углов гибкого штампа с расчетом

Подробнее рассмотрим копирование эшелеты. На рис. 5 изображена дифракционная решетка, нарезанная на металле, которая служила мастер-штампом. Данная решетка представляла собой одну или две решетки, пересекающиеся под углом 90° с частотой 1800 лин/мм в обоих вариантах. Следовательно, структура имела размер элементов менее 500×500 нм, имеющих фиксированные углы и достаточно большую высоту рельефа (70 нм).

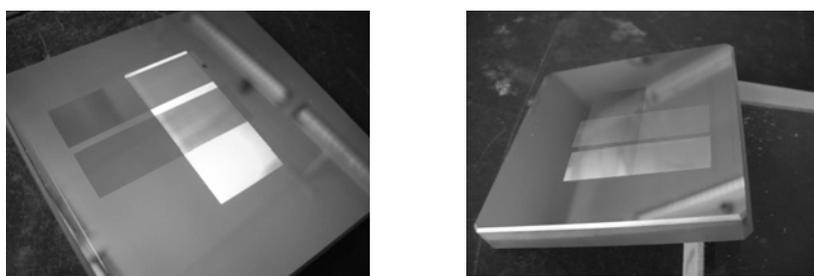


Рис. 5. Фотография дифракционной решетки – оригинала

При копировании поверхность шаблона должна быть чистой и свободной от загрязнений. Для этого необходимо провести предварительное очищение шаблона, поместив его в состав растворителя на несколько минут.

Следующим этапом является нанесение силикона на шаблон (оригинал), изготавливаем штамп с учетом вышеуказанных особенностей. Под микроскопом видно, что силикон полностью затекает во все мелкие элементы оригинала, и форма получаемого штампа передается с хорошей точностью, что подтверждается полученными копиями дифракционной решетки рис. 6. Далее на стекло наносится УФ-полимеризуемая композиция, сверху кладем штамп и помещаем образцы в установку рис. 3, вакуумируем в течение 40 минут и после напуска аргона экспонируем 3 минуты.



Рис. 6. Копия дифракционной решетки

Проверка соответствия передачи формы мастер-штампа при копировании проводилась как при получении фотографии рельефа поверхности с использованием атомно-силового микроскопа, так и по наблюдению дифракционной картины мастер-штампа и копии. Частота 1800 лин/мм является практически предельной для механического способа нарезки. Поэтому необходимо учесть возможные дефекты поверхности мастер-штампа, которые передаются на копию. Так, отмечено, что при изготовлении эшелет с высокой частотой происходит «заваливание» зубьев, в связи с выдавливанием материала. На рис. 7 показана форма поверхности копии, измеренная на сканирующем туннельном микроскопе.

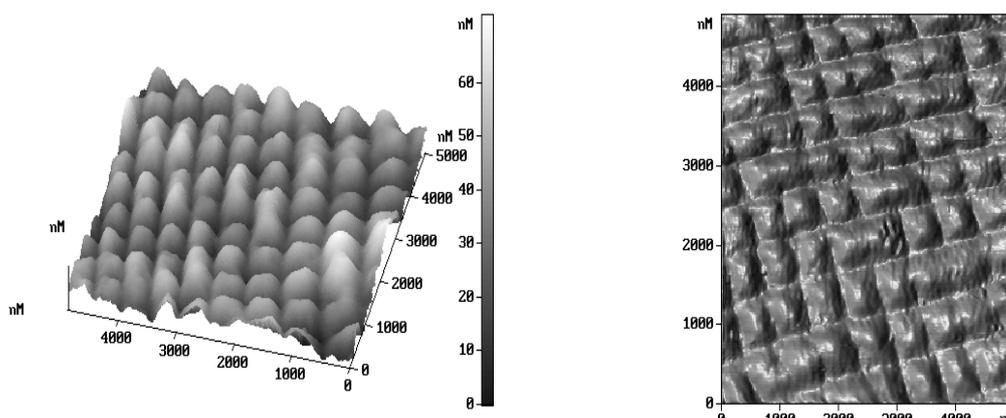


Рис. 7. Форма поверхности копии

При анализе данной фотографии необходимо учитывать также и возможные неточности туннельного микроскопа при сканировании неровной поверхности, поэтому нами были также проведены эксперименты по сравнению оригинала и копии по получаемой дифракционной картине. На рис. 8 представлена дифракционная картина, полученная на данных решетках.



Рис. 8. Дифракционная картина, полученная на решетках: при отражении света от матрицы – оригинала (слева); при отражении света от копии, полученной рассмотренным методом (справа)

Сравнение фотографий показывает, что дифракционные картины подобны, и копия является эшелетой с ориентированным зубом, также как и мастер-штамп. Следовательно, данный метод пригоден для копирования нанометровых структур. Однако необходимо провести дальнейшие эксперименты для определения точной величины погрешностей при копировании нанометровых объектов.

Выводы

С использованием предложенного метода было выполнено копирование френелевской линзы. Сравнение копии с оригиналом показывает, что изменение формы в результате копирования составляет менее 1%. Копирование эшелеты подтверждает возможность использования метода для нанометровых структур. Таким образом, экспериментальная реализация метода показала его пригодность для формирования микро- и наноэлементов фотоники с высокой точностью при возможности их массового тиражирования. Можно сказать, что для большинства задач по созданию наноструктур наноимпринт является технологией, обеспечивающей высокое качество и гибкую адаптацию к процессу при малых вложениях.

Литература

1. Byron D. Gates // *Materials today*. – February 2005. – P. 44–49..
2. Otto M., Bender M. Reproducibility and homogeneity in step and repeat UV-nanoimprint lithography // *Microelectronic Engineering*. – № 73–74 (2004). – P.152–156
3. Компания «Пента». Официальный сайт [Электронный ресурс]. – Режим доступа: <http://www.penta-91.ru>
4. Смирнова Т.В., Бурункова Ю.Э., Денисюк И.Ю. Измерение усадок УФ-отверждаемых композиций на основе акрилатов и диакрилатов. // *Оптический журнал*. – 2006. – № 5. – С 57–61.
5. Andrejewska E. Photopolymerization kinetics of multifunctional monomers // *Progress in polymer science*. – 2001. – №26. – P. 605–665
6. Gates D. Nanofabrication with molds and stamps // *Materials today*. – 2005, February. – P. 44–49.
7. Sotomayor Torres C.M., Zankovych S. Nanoimprint lithography: an alternative nanofabrication approach // *Materials Science and Engineering*. – 2003. – №23. – P. 23–31.

РЕЗОНАНСНЫЙ ПЕРЕНОС ЭНЕРГИИ ФОТОВОЗБУЖДЕНИЯ В КВАНТОВЫХ ТОЧКАХ

С.Ю. Кручинин

Научный руководитель – д.ф.-м.н., ст.н.с. А.В. Федоров

Исследован резонансный перенос энергии между квантовыми точками вследствие электростатического взаимодействия. Вычислена вероятность этого процесса для прямозонного полупроводника в предположении о том, что взаимодействие между электронными подсистемами точек описывается экранированным кулоновским потенциалом. При этом учитывались все существенные члены мультипольного разложения, что позволило адекватно рассмотреть случай запрещенных переходов. Показано, что вероятность переноса в этом случае существенно выше, чем в случае молекулярных систем, поскольку мультипольные взаимодействия вносят вклад в дипольные межзонные переходы.

1. Введение

Изучение безызлучательного переноса энергии между парой пространственно разнесенных квантовых объектов (атомы, молекулы, агрегаты, примесные центры, наноструктуры) является важной задачей физики конденсированного состояния [1], в том числе физики низкоразмерных систем [2, 3]. В последние годы исследования этого эффекта интенсивно развивались в системах с квантовыми точками (нанокристаллами) [4–11]. Интерес к таким системам обусловлен прежде всего перспективами их использования в различных приложениях, например, при разработке люминесцентных меток и биосенсоров, применяемых для визуализации биологических объектов [12, 13] и экологического контроля окружающей среды, при создании низкопороговых лазеров [14], элементной базы нанoeлектроники, квантовых компьютеров [15–17] и клеточных автоматов [18]. Кроме этого, нанокристаллы являются хорошим модельным объектом для детального изучения безызлучательного переноса энергии. Благодаря эффекту размерного квантования [19] можно осуществить резонанс между любыми электронными уровнями квантовой точки донора энергии и квантовой точки акцептора путем подбора соответствующих размеров нанокристаллов. Таким образом, можно исследовать зависимость эффективности переноса энергии от свойств электронных состояний, участвующих в этом процессе.

Диполь-дипольное приближение использовалось для теоретического описания переноса энергии [8, 9, 23] и интерпретации экспериментальных данных [4, 10, 24, 25] во многих современных работах, посвященных квантовым точкам. Очевидно, что если расстояние между квантовыми точками значительно больше их размеров, то диполь-дипольное приближение адекватно описывает перенос энергии фотовозбуждений. Однако возникает вопрос о его применимости для часто реализующегося на практике случая, когда квантовые точки находятся на расстоянии, сопоставимом с их размерами. На первый взгляд, в этой ситуации следует учитывать мультипольные взаимодействия более высокого порядка. Наиболее просто это можно сделать, если предположить, что взаимодействие электронов донора и акцептора описывается кулоновским потенциалом, и учитывать его диэлектрическое экранирование с помощью некоторой эффективной диэлектрической проницаемости, не зависящей от координат. Данный подход использовался в ряде теоретических работ [11, 15]. Например, в [15] механизм Ферстера исследовался на предмет возможности создания запутанных состояний в квантовых точках. Путем численных расчетов показано, что диполь-дипольное приближение адекватно даже на малых расстояниях для квантовых точек кубической формы, а для точек в форме параллелепипедов возникают заметные отличия. В работе [11] в рамках приближения сильной связи рассмотрен перенос в сферических квантовых точках прямозонных и непрямозонных полупроводников. Показано, что для прямозонных полупроводников на малых расстояниях диполь-дипольное приближение

адекватно, и можно пренебречь членами мультипольного разложения. Установлено, что для непрямозонных полупроводников мультипольные члены играют более существенную роль, однако диполь-дипольное взаимодействие все равно остается доминирующим. В [11] также был предложен способ вычисления константы, описывающей диэлектрическое экранирование кулоновского взаимодействия в системах с нанокристаллами.

Перечисленные результаты противоречат оценке отношения вероятностей диполь-квадрупольного и диполь-дипольного взаимодействий [22], согласно которой при контакте квантовых точек относительный вклад диполь-квадрупольного взаимодействия в вероятность переноса энергии может составлять величину $\sim 20\%$. Анализ этого противоречия является одной из целей данной работы. Он проводится на примере взаимодействия двух квантовых точек сферической формы. Это позволило получить ряд интересных результатов: правила отбора межзонных переходов и относительно простое выражение матричного элемента. Показано, что в системах с квантовыми точками, в отличие от молекулярных или атомных [1, 22], учет мультипольных взаимодействий приводит к значительно большим величинам вероятности переноса в случае дипольно-запрещенных переходов. С другой стороны, когда переходы в доноре и акцепторе являются дипольно-разрешенными, выражение матричного элемента точно переходит в формулу, получаемую в диполь-дипольном приближении.

Работа построена следующим образом. В разделе 2 вычислен матричный элемент взаимодействия, подробно рассматривается его зависимость от расстояния между донором и акцептором в случае разрешенных и запрещенных переходов, приводятся выражения, определяющие правила отбора для переходов вследствие переноса энергии. Затем обсуждается обобщение полученного выражения для матричного элемента на случай квантовых точек с конечной высотой потенциального барьера. В разделе 3 приводится выражение скорости безызлучательного переноса между квантовыми точками, анализируется его зависимость от температуры.

2. Вероятность переноса энергии

Рассмотрим квантовую точку-донор и квантовую точку-акцептор сферической формы с радиусами R_D и R_A ($R_D < R_A$), расположенные на расстоянии \mathbf{r} друг от друга в диэлектрической матрице (рис. 1).

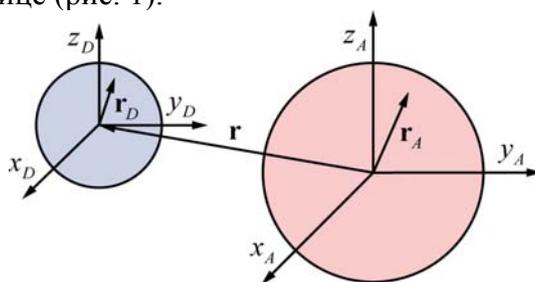


Рис. 1. Системы координат, связанные с донором и акцептором

Будем считать, что в начальном состоянии системы донор–акцептор в зоне проводимости донора имеется электрон, а валентная зона акцептора полностью заполнена. В результате кулоновского взаимодействия, описываемого потенциалом

$$V(\mathbf{r}, \mathbf{r}_D, \mathbf{r}_A) = \frac{e^2}{\varepsilon |\mathbf{r} + \mathbf{r}_D - \mathbf{r}_A|}, \quad (1)$$

система переходит в конечное состояние с электроном в валентной зоне донора и электроном в зоне проводимости акцептора. В выражении (1) \mathbf{r}_D и \mathbf{r}_A – векторы, отсчитываемые от центров соответствующих квантовых точек, \mathbf{r} – вектор, направленный от центра акцептора к центру донора, экранирование кулоновского потенциала учитывается с помощью эффективной диэлектрической проницаемости ε [11], которая выражается через высокочастотные проницаемости донора ε_D , акцептора ε_A и матрицы ε_M

$$\varepsilon = \frac{(\varepsilon_D + 2\varepsilon_M)(\varepsilon_A + 2\varepsilon_M)}{9\varepsilon_M}. \quad (2)$$

Матричный элемент межточечного кулоновского взаимодействия электронной подсистемы донора и акцептора, соответствующий процессу безызлучательного переноса энергии, можно представить следующим образом

$$M_{DA} \equiv \langle i', f' | H_{DA} | i, f \rangle = \frac{e^2}{\varepsilon} \int d^3r_D d^3r_A \frac{\Psi_{Dv i'}^*(\mathbf{r}_D) \Psi_{Acf}^*(\mathbf{r}_A) \Psi_{Dci}(\mathbf{r}_D) \Psi_{Afv}(\mathbf{r}_A)}{|\mathbf{r} + \mathbf{r}_D - \mathbf{r}_A|}, \quad (3)$$

где $\Psi_{\alpha jt}(\mathbf{r}_\alpha)$ – волновая функция электрона донора $\alpha = D$ (акцептора $\alpha = A$), находящегося в зоне $j = c, v$ в состоянии с квантовыми числами $t = i, i'$ (донор) или $t = f, f'$ (акцептор). Функции $\Psi_{\alpha jt}(\mathbf{r}_\alpha)$ имеют, в общем случае, вид линейных комбинаций произведений блоховской амплитуды u_j и огибающих волновых функций $\psi_{\alpha jt}$.

В работах [15, 26] показано, что обменное кулоновское взаимодействие мало, поэтому оно не учитывалось при вычислении (3).

При описании электронной подсистемы будем считать, что реализуется режим сильного конфайнмента, и пренебрегать кулоновским взаимодействием между электроном и дыркой, которые находятся внутри одной квантовой точки, так как его учет приводит к несущественному для данной задачи сдвигу энергетических уровней и изменению волновых функций.

Для вычисления матричного элемента (3) воспользуемся двухзонной моделью полупроводника. В рамках метода эффективной массы [27] волновая функция носителя заряда $\Psi_{\alpha jt}(\mathbf{r}) = u_{\alpha j}(\mathbf{r})\psi_{\alpha jt}(\mathbf{r})$ может быть представлена в виде произведения блоховской амплитуды $u_{\alpha j}$ и медленной огибающей $\psi_{\alpha jt}$. В режиме сильного конфайнмента [19] огибающая волновая функция носителей заряда в сферических квантовых точках для случая бесконечно высоких потенциальных стенок имеет вид

$$\psi_{\alpha jt}(\mathbf{r}_\alpha) = \sqrt{\frac{2}{R_\alpha^3} \frac{j_{l_\alpha}(\xi_{n_\alpha l_\alpha} r_\alpha / R_\alpha)}{j_{l_\alpha+1}(\xi_{n_\alpha l_\alpha})}} Y_{l_\alpha m_\alpha}(\theta_\alpha, \varphi_\alpha), \quad (4)$$

где $t = \{nlm\}$, n , l , m — главное квантовое число, угловой момент и его проекция, R – радиус квантовой точки, $j_l(x)$ – сферические функции Бесселя, Y_{lm} – сферическая гармоника, ξ_{nl} – n -й корень выражения $j_l(x) = 0$. Следует отметить, что использование бесконечно высоких потенциальных стенок для описания полупроводниковых квантовых точек в диэлектрических матрицах является хорошим приближением, поскольку в системе полупроводник-диэлектрик потенциальные ямы для электронов и дырок, как правило, очень глубокие. Если воспользоваться Фурье-разложением потенциала

$$\frac{1}{|\mathbf{r} + \mathbf{r}_D - \mathbf{r}_A|} = \frac{1}{2\pi^2} \int d^3q \frac{e^{iq(\mathbf{r} + \mathbf{r}_D - \mathbf{r}_A)}}{q^2}, \quad (5)$$

то матричный элемент (3) будет иметь вид

$$M_{DA} = \frac{e^2}{\varepsilon} \frac{1}{2\pi^2} \int d^3q \frac{e^{i\mathbf{q}\mathbf{r}}}{q^2} S_{ii'}^{(D)}(\mathbf{q}) S_{ff'}^{(A)*}(\mathbf{q}), \quad (6)$$

где

$$S_{ii'}^{(\alpha)}(\mathbf{q}) = \int d^3r_\alpha \Psi_{\alpha i'}^*(\mathbf{r}_\alpha) e^{i\mathbf{q}\mathbf{r}_\alpha} \Psi_{\alpha i}(\mathbf{r}_\alpha). \quad (7)$$

Представим \mathbf{r}_α в виде суммы радиус-вектора элементарной ячейки \mathbf{r}_{k_α} и радиус-вектора электрона относительно нее \mathbf{r}'_α . Воспользуемся тем, что блоховские функции периодичны с периодом решетки, а огибающие медленно меняются в пределах элементарной ячейки:

$$\Psi_{\alpha j}(\mathbf{r}_{k_\alpha} + \mathbf{r}'_\alpha) \approx \Psi_{\alpha j}(\mathbf{r}_{k_\alpha}), \quad u_{\alpha j}(\mathbf{r}_{k_\alpha} + \mathbf{r}'_\alpha) = u_{\alpha j}(\mathbf{r}'_\alpha).$$

Перейдем в (7) от интегрирования по объему квантовых точек V_α к сумме интегралов по объемам элементарных ячеек кристалла Ω_α . В результате получим

$$S_{ii'}^{(\alpha)}(\mathbf{q}) = \sigma_{ii'}^{(\alpha)} \frac{1}{\Omega_\alpha} \int d^3r'_\alpha u_{\alpha i'}^*(\mathbf{r}'_\alpha) e^{i\mathbf{q}\mathbf{r}'_\alpha} u_{\alpha i}(\mathbf{r}'_\alpha), \quad (8)$$

$$\sigma_{ii'}^{(\alpha)} = \Omega_\alpha \sum_{k_\alpha} \Psi_{\alpha i'}^*(\mathbf{r}_{k_\alpha}) e^{i\mathbf{q}\mathbf{r}_{k_\alpha}} \Psi_{\alpha i}(\mathbf{r}_{k_\alpha}). \quad (9)$$

Заменяя под интегралом $e^{i\mathbf{q}\mathbf{r}'_\alpha}$ на $\cos \mathbf{q}\mathbf{r}'_\alpha + i \sin \mathbf{q}\mathbf{r}'_\alpha$, разложим подынтегральное выражение на два различных по четности слагаемых. В силу того, что произведение блоховских функций зоны проводимости и валентной зоны нечетно, отличным от нуля будет интеграл с $i \sin \mathbf{q}\mathbf{r}'_\alpha$. Раскладывая синус в ряд и ограничиваясь первым ненулевым членом, получим

$$S_{ii'}^{(\alpha)}(\mathbf{q}) = i \sigma_{ii'}^{(\alpha)}(\mathbf{q}\mathbf{r}_{vc}^{(\alpha)}), \quad (10)$$

где

$$\mathbf{r}_{vc}^{(\alpha)} = \frac{1}{\Omega_\alpha} \int_{\Omega_\alpha} d^3r'_\alpha u_{\alpha i'}^*(\mathbf{r}'_\alpha) \mathbf{r}'_\alpha u_{\alpha i}(\mathbf{r}'_\alpha), \quad (11)$$

– матричный элемент координаты на блоховских функциях. Его модуль можно выразить через матричный элемент импульса и связанный с ним параметр Кейна $P^{(\alpha)} = \hbar^2 / m_0 \langle S | \partial / \partial z | Z \rangle$ [28]

$$|\mathbf{r}_{vc}^{(\alpha)}| = \frac{P^{(\alpha)}}{E_g^{(\alpha)}},$$

m_0 – масса свободного электрона, $E_g^{(\alpha)}$ – ширина запрещенной зоны в объемном полупроводнике. Переходя в (9) от суммирования по элементарным ячейкам к интегрированию по объему нанокристалла, получаем

$$\sigma_{ii'}^{(\alpha)} = \int d^3r_\alpha \Psi_{\alpha i'}^*(\mathbf{r}_\alpha) e^{i\mathbf{q}\mathbf{r}_\alpha} \Psi_{\alpha i}(\mathbf{r}_\alpha). \quad (12)$$

Используя (10) и (12), матричный элемент (3) можно представить следующим образом

$$M_{DA} = \frac{e^2}{\varepsilon} \frac{1}{2\pi^2} \int d^3q \frac{e^{i\mathbf{q}\mathbf{r}}}{q^2} (\mathbf{q}\mathbf{r}_{vc}^{(D)}) (\mathbf{q}\mathbf{r}_{cv}^{(A)}) \sigma_{ii'}^{(D)} \sigma_{ff'}^{(A)*}. \quad (13)$$

Для дальнейшего упрощения (12) воспользуемся известным разложением плоской волны по сферическим гармоникам. Сферическая симметрия позволяет устранить зависимость $\sigma_{ii'}^{(\alpha)}$ от угловых координат вектора \mathbf{q} , полагая его сонаправленным с осями z_α систем координат, связанных с соответствующими квантовыми точками. С помощью формулы, выражающей интеграл от трех сферических функций через коэффициенты Клебша–Гордана [29], преобразуем $\sigma_{ii'}^{(\alpha)}$ к следующему виду:

$$\sigma_{ii'}^{(\alpha)}(q) = \frac{2}{R_\alpha^3} \sqrt{\frac{2l_\alpha + 1}{2l'_\alpha + 1}} \sum_{l=0}^{\infty} i^l (2l+1) C_{l_\alpha 0, l 0}^{l'_\alpha 0} C_{l_\alpha m_\alpha, l 0}^{l'_\alpha m'_\alpha} \times \int_0^{R_\alpha} dr_\alpha r_\alpha^2 \frac{j_{l_\alpha}(k_{n_\alpha l_\alpha} r_\alpha) j_{l'_\alpha}(k_{n_\alpha l'_\alpha} r_\alpha)}{j_{l_\alpha+1}(\xi_{n_\alpha l_\alpha}) j_{l'_\alpha+1}(\xi_{n_\alpha l'_\alpha})} j_l(qr_\alpha), \quad (14)$$

где $k_{n_\alpha l_\alpha} = \xi_{n_\alpha l_\alpha} / R_\alpha$. Произведения коэффициентов Клебша–Гордана

$$C_{l_\alpha 0, l 0}^{l'_\alpha 0} C_{l_\alpha m_\alpha, l 0}^{l'_\alpha m'_\alpha}$$

определяют правила отбора для переходов носителей заряда в процессе переноса энергии:

$$|l_\alpha - l| \leq l'_\alpha \leq l_\alpha + l, \quad (15)$$

$$l_\alpha + l'_\alpha + l_\alpha \text{ — четное число,} \quad (16)$$

$$m_\alpha = m'_\alpha. \quad (17)$$

Помимо этого, существуют нетривиальные случаи обращения коэффициентов Клебша–Гордана в ноль [29], поэтому не представляется возможным записать общее выражение, определяющее правила отбора.

Теперь в (13) можно вычислить интеграл по угловым переменным, выбирая систему координат так, что ось z совпадает с вектором \mathbf{r} . В результате приходим к следующему выражению матричного элемента (13)

$$M_{DA} = \frac{e^2}{\varepsilon r^3} \left[I^{(1)} \mathbf{r}_{vc}^{(D)} \mathbf{r}_{cv}^{(A)} - I^{(2)} 3(\mathbf{n}_r \mathbf{r}_{vc}^{(D)}) (\mathbf{n}_r \mathbf{r}_{cv}^{(A)}) \right], \quad (18)$$

где

$$I^{(l)} = \left(\frac{1}{3} \right)^{l-1} \frac{2}{\pi} \int_0^\infty dx x^l j_l(x) \sigma_{ii'}^{(D)}(x/r) \sigma_{ff'}^{(A)*}(x/r), \quad l=1,2, \quad x=qr, \quad (19)$$

\mathbf{n}_r – единичный вектор в направлении \mathbf{r} . Анализ выражений (19) позволяет получить дополнительные правила отбора для переходов носителей заряда в процессе переноса энергии. Величины $I^{(l)}$ отличны от нуля, когда $l_D = l'_D$ и $l_A = l'_A$, а также когда $l_D + l'_D$ и $l_A + l'_A$ обладают разной четностью.

В дипольном приближении матричный элемент (18) имеет вид

$$M_{DA}^{(d-d)} = \frac{e^2}{\varepsilon r^3} \delta_{ii'} \delta_{ff'} \left[\mathbf{r}_{vc}^{(D)} \mathbf{r}_{cv}^{(A)} - 3(\mathbf{n}_r \mathbf{r}_{vc}^{(D)}) (\mathbf{n}_r \mathbf{r}_{cv}^{(A)}) \right], \quad (20)$$

где символы Кронекера показывают, что перенос энергии возможен лишь между теми состояниями, между которыми разрешен межзонный переход в дипольном приближении. Видно, что выражение (18) формально подобно (20). Однако, в отличие от последнего, в (18) фигурируют амплитуды $I^{(l)}$. Кроме этого, из (18) следует, что возможен перенос энергии между состояниями, для которых межзонный переход запрещен в дипольном приближении ($i \neq i'$ и $f \neq f'$).

Анализ показывает, что для дипольно-разрешенных межзонных переходов ($i = i'$ и $f = f'$) выражение (18) полностью совпадает с (20) и закон r^{-3} выполняется для любых расстояний r между квантовыми точками. Иначе говоря, в этом случае $I^{(1)} = I^{(2)} = 1$. Этот результат кажется неожиданным, поскольку согласно оценкам, выполненным в работе [22], отношение матричного элемента диполь-квадрупольного взаимодействия $M_{DA}^{(d-q)}$ к матричному элементу диполь-дипольного взаимодействия $M_{DA}^{(d-d)}$ по порядку величины должно быть равно $\beta = R_\alpha / r$. В нашем случае при контакте донора и акцептора этот параметр $\beta \cong 1/2$. По всей видимости, совпадение

матричного элемента (18) с (20) связано с высокой симметрией рассматриваемой задачи. Данный вывод подтверждается прямыми численными расчетами, выполненными для сферических [11] и кубических [15] квантовых точек. В то же время в работе [15] было показано, что для квантовых точек в форме прямоугольных параллелепипедов имеется заметное отклонение от диполь-дипольного приближения для малых расстояний между донором и акцептором.

Отметим, что в используемом приближении угловая зависимость матричного элемента совпадает с той, что имеет место в случае диполь-дипольного взаимодействия. Это обусловлено симметрией задачи, благодаря которой вектор \mathbf{q} не имеет выделенного направления в системах координат, связанных с центрами квантовых точек, и при разложении плоской волны в (12) можно ограничиться сферическими гармониками с нулевыми проекциями углового момента.

Если квантовые точки изготовлены из одного материала, квадрат модуля матричного элемента (18), усредненный по направлениям дипольных моментов межзонных переходов, будет равен

$$\overline{|M_{DA}|^2} = \frac{1}{3} \frac{e^4}{\varepsilon^2 r^6} \left(\frac{P}{E_g} \right)^4 \left\{ |I^{(1)}|^2 - (I^{(1)*} I^{(2)} + I^{(1)} I^{(2)*}) + 3 |I^{(2)}|^2 \right\}. \quad (21)$$

В диполь-дипольном приближении или для дипольно-разрешенных межзонных переходов он имеет вид

$$\overline{|M_{DA}^{(dd)}|^2} = \frac{2}{3} \frac{e^4}{\varepsilon^2 r^6} \left(\frac{P}{E_g} \right)^4 \delta_{n_D n'_D} \delta_{n_A n'_A} \delta_{l_D l'_D} \delta_{l_A l'_A} \delta_{m_D m'_D} \delta_{m_A m'_A}. \quad (22)$$

Проанализируем матричный элемент переноса энергии для нескольких частных случаев, а именно, для нижайших по энергии переходов, для которых он отличен от нуля. Будем считать, что в начальном состоянии электрон и дырка донора находятся в нижайших состояниях $i = i' = \{100\}$. Эта ситуация представляет особый интерес в том случае, когда скорость внутризонной релаксации носителей заряда существенно превышает скорость переноса энергии. Предположим, что квантовые точки донор и акцептор изготовлены из одного и того же материала. Для определенности рассмотрим кубическую модификацию CdSe [30]: $m_c^{(\alpha)} = 0.11m_0$, $m_v^{(\alpha)} = 1.14m_0$, $E_g^{(\alpha)}(293\text{ K}) = 1.736\text{ эВ}$, $P = 1.48 \cdot 10^{-19}\text{ см}^3\text{ г с}^{-2}$, $\varepsilon_\alpha = 5.8$. В качестве матрицы выберем SiO_2 с высокочастотной диэлектрической проницаемостью $\varepsilon_M = 2.13$ [31]. При этом эффективная диэлектрическая проницаемость (2) равна $\varepsilon = 5.28$. Очевидно, что скорость прямого переноса энергии будет максимальна, когда энергия перехода в доноре будет равна энергии перехода в акцепторе. В рассматриваемом случае для этого необходимо, чтобы радиусы квантовых точек R_D и R_A удовлетворяли соотношению

$$\frac{R_D}{R_A} = \pi \sqrt{\frac{m_c + m_v}{m_c \xi_{nl}^2 + m_v \xi_{n'l'}^2}}, \quad (23)$$

где n, l и n', l' – квантовые числа состояний в валентной зоне и зоне проводимости акцептора.

На рис. 4 представлена зависимость $\overline{|M_{DA}|^2}$ от расстояния между центрами квантовых точек для различных переходов в акцепторе: дипольно разрешенные переходы $Av\{nlm\} \rightarrow Ac\{nlm\}$ и дипольно запрещенные переходы $Av\{110\} \rightarrow Ac\{120\}$, $Av\{120\} \rightarrow Ac\{110\}$, $Av\{100\} \rightarrow Ac\{110\}$, $Av\{110\} \rightarrow Ac\{100\}$.

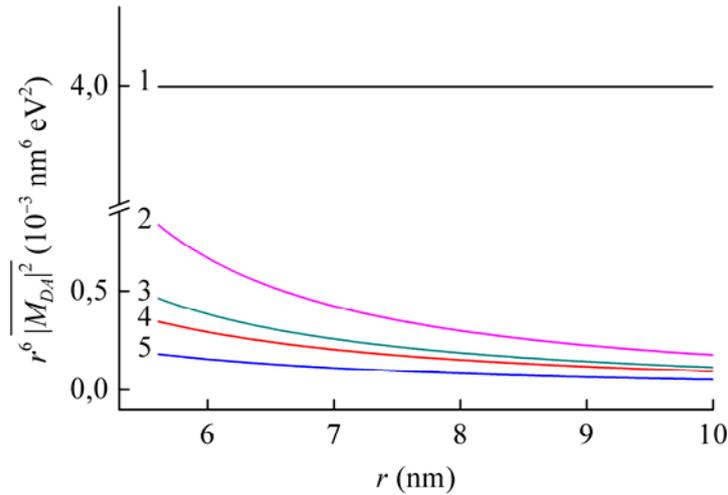


Рис. 4. Зависимость квадрата модуля матричного элемента взаимодействия от расстояния между центрами донора и акцептора. 1 — $Av\{nlm\} \rightarrow Ac\{nlm\}$,
 2 — $Av\{110\} \rightarrow Ac\{120\}$, 3 — $Av\{120\} \rightarrow Ac\{110\}$,
 4 — $Av\{100\} \rightarrow Ac\{110\}$, 5 — $Av\{110\} \rightarrow Ac\{100\}$

В расчетах предполагалось, что радиус донора $R_D = 2$ нм, а радиус акцептора подбирался в согласии с (23) так, чтобы соответствующие переходы попадали в резонанс. Из рис. 4 следует, что относительная величина квадрата модуля матричного элемента для дипольно запрещенных переходов может достигать $\sim 20\%$ от соответствующей величины для дипольно разрешенных переходов при малых расстояниях между квантовыми точками. Поскольку перенос в дипольно запрещенные состояния акцептора обусловлен мультипольным взаимодействием, начиная с квадрупольного, то полученный результат согласуется с приведенной выше оценкой, которая для рассматриваемого случая дает $M_{DA}^{(dq)}/M_{DA}^{(dd)} = \beta^2 = (R_A/r)^2 \leq 25\%$.

3. Вероятность переноса энергии

Используя (21), можно вычислить скорость безызлучательного переноса энергии из некоторого фиксированного состояния донора во все возможные состояния акцептора γ_{DA} . Поскольку взаимодействие носителей донора и акцептора является достаточно слабым, то для расчета вероятности переноса энергии электрон-дырочной пары донора в состоянии $|i_D, i'_D\rangle$ к электрон-дырочной паре акцептора в состоянии $|f_A, f'_A\rangle$ можно ограничиться первым порядком теории возмущений. Тогда скорость переноса энергии можно записать следующим образом:

$$\gamma_{DA} = \frac{2}{\hbar^2} \overline{|M_{DA}|^2} \frac{\Gamma_{DA}}{\Gamma_{DA}^2 + \Delta_{DA}^2}, \quad (25)$$

где

$$\Delta_{DA} = \left(E_{c,i}^{(D)} + E_{v,i'}^{(D)} + E_g^{(D)} - E_{c,f'}^{(A)} - E_{v,f}^{(A)} - E_g^{(A)} \right) / \hbar, \quad (26)$$

частотная расстройка между уровнями электрон-дырочных пар, участвующих в переходе, $E_{j,t}^{(\alpha)} = \hbar^2 \xi_{n_\alpha l_\alpha}^2 / (2m_j^{(\alpha)} R_\alpha^2)$, Γ_{DA} – скорость релаксации когерентности между начальными конечными состояниями, суммирование проводится по всем возможным состояниям квантовой точки акцептора. В случае дипольно-разрешенных переходов скорость переноса энергии из донора, электрон и дырка которого находится в состояниях с квантовыми числами n_D, l_D, m_D , во все возможные состояния акцептора равна

$$\gamma_{DA}^{(d-d)} = \frac{4e^4}{3\hbar^2 \varepsilon^2 r^6} \left(\frac{P}{E_g} \right)^4 \sum_{n_A l_A} (2l_A + 1) \frac{\Gamma_{DA}}{\Gamma_{DA}^2 + \Delta_{DA}^2},$$

Определим Γ_{DA} как сумму скоростей дефазировки межзонных переходов в доноре и акцепторе:

$$\Gamma_{DA} = \gamma_{i_D i_D} + \gamma_{f_A f_A}, \quad (28)$$

где $\gamma_{x' x_a} = (\gamma_{x'_a} + \gamma_{x_a}) / 2 + \bar{\gamma}_{x'_a x_a}$, γ_{x_a} – обратное время жизни состояния x в доноре или акцепторе, $\bar{\gamma}_{x'_a x_a}$ – чистая дефазировка перехода. Отсюда следует, что скорость переноса энергии существенно зависит от температуры, что согласуется с экспериментальными данными [4].

Конечная ширина переходов в процессе переноса энергии приводит к ряду важных следствий. Прежде всего, оказывается, что для переноса энергии от электронной подсистемы донора к электронной подсистеме акцептора не требуется точного резонанса между соответствующими электрон-дырочными состояниями. Проиллюстрируем это утверждение на примере перехода между дипольно разрешенными состояниями CdSe квантовых точек донора и акцептора $Dc\{100\}$, $Dv\{100\} \rightarrow Av\{11m\}$, $Ac\{11m\}$. На рис. 5 представлена зависимость вероятности переноса энергии γ_{DA} от энергетической расстройки между уровнями электрон-дырочных пар, участвующих в переходе, для комнатной и гелиевой температур.

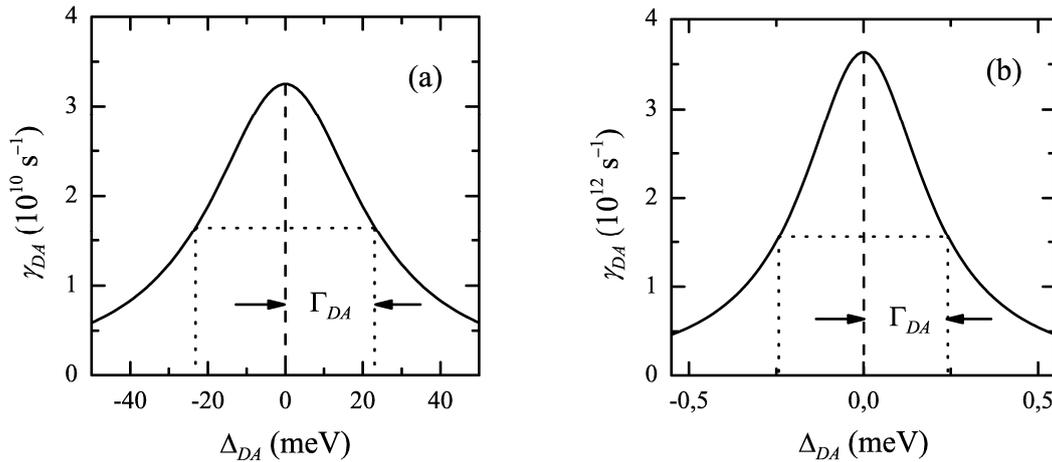


Рис. 5. Вероятность переноса энергии γ_{DA} для CdSe квантовых точек донора и акцептора в зависимости от энергетической отстройки Δ_{DA} от резонанса между дипольно разрешенными состояниями электрон-дырочных пар, участвующих в процессе (а) при $T = 300$ К, (б) при $T = 4$ К. Расстояние между квантовыми точками $r = 5$ нм, Γ_{DA} – скорость дефазировки перехода (28) в энергетических единицах

В расчете использовалось феноменологическое выражение для скорости дефазировки межзонных переходов, часто применяемое при анализе экспериментальных данных [33, 34], полученных в системах с квантовыми точками,

$$\gamma_{cv} = \gamma_0 + aT + b\bar{n}_{LO}(T), \quad (29)$$

где $\bar{n}_{LO}(T) = 1 / [\exp(\hbar\omega_{LO} / k_B T) - 1]$, ω_{LO} – частота продольного оптического фонона в доноре или акцепторе, k_B – постоянная Больцмана, T – температура. В (29) первое слагаемое, γ_0 , интерпретируется как скорость дефазировки, обусловленной взаимодействием с термостатом (безызлучательные и излучательные межзонные переходы), второе и третье слагаемые описывают взаимодействие с акустическими и

оптическими фононами ($a = 1.5 \cdot 10^{10} \text{ c}^{-1} / \text{K}$ [33, 34], $b = 2.3 \cdot 10^{13} \text{ c}^{-1}$ [33], $\hbar\omega_{LO} = 26 \text{ мэВ}$). Что касается величины γ_0 , то для донора, находящегося в нижайшем по энергии состоянии электрон-дырочных пар, было выбрано значение $7.7 \cdot 10^7 \text{ c}^{-1}$, что соответствует квантовому выходу 50%, поскольку радиационная ширина перехода составляет $3.85 \cdot 10^7 \text{ c}^{-1}$ [35]. Для акцептора, находящегося в высокоэнергетическом состоянии, необходимо учесть вклад в γ_0 от внутризонной релаксации электрон-дырочных пар. Согласно работе [36], скорость внутризонной релаксации при переходах между нижайшими по энергии состояниями электрон-дырочных пар в CdSe квантовых точках может меняться от 10^{11} c^{-1} до $3.3 \cdot 10^{12} \text{ c}^{-1}$ при уменьшении радиуса квантовой точки от 5.6 нм до 2 нм. В приведенном на рис. 5 расчете вклад в γ_0 от внутризонной релаксации акцептора был выбран, для определенности, равным $2 \cdot 10^{11} \text{ c}^{-1}$, предполагалось, что расстояние между донором и акцептором равно $r = 5 \text{ нм}$. Из рис. 5 видно, что в пределах ширины Γ_{DA} возможен эффективный прямой перенос как с понижением, так и с повышением энергии. При уменьшении температуры (рис. 5, б) вероятность переноса энергии заметно возрастает, а ширина перехода уменьшается. При комнатной температуре (рис. 5, а) может реализоваться такая ситуация, когда в спектральную область $-\Gamma_{DA} \leq \Delta_{DA} \leq \Gamma_{DA}$ попадут несколько уровней электрон-дырочных пар акцептора. В этом случае будет иметь место эффективный перенос энергии одновременно на все эти уровни. Следует подчеркнуть, что применение в выполненных расчетах выражения (29) с приведенными выше значениями параметров γ_0 , a , b , $\hbar\omega_{LO}$ эквивалентно использованию экспериментальных данных.

4. Заключение

В работе исследовано явление резонансного переноса энергии в системе двух квантовых точек. Для точек сферической формы получено выражение вероятности переноса в предположении о том, что взаимодействие между носителями донора и акцептора описывается кулоновским потенциалом. Это позволило адекватно рассмотреть случаи, когда квантовые точки находятся на расстояниях, сопоставимых с их размерами, и когда межзонный переход в акцепторе является дипольно-запрещенным. Вычисление матричного элемента проводилось в рамках двухзонного приближения, что позволило прийти к достаточно простым выражениям для вероятности переноса энергии. Показано, что диполь-дипольное приближение адекватно даже на малых расстояниях между нанокристаллами, а вероятность переноса с переходом акцептора в дипольно-запрещенное состояние существенно выше, чем в атомных или молекулярных системах. Так, согласно [1, 22], в молекулярных системах отношение вероятностей диполь-дипольного и диполь-квадрупольного переноса составляет $W_{dq} / W_{dd} \sim 10^{-7}$. Показано, что в системах с квантовыми точками учет мультипольных взаимодействий приводит к значительно большим величинам вероятности переноса в случае дипольно-запрещенных переходов. В частности, на расстояниях менее 2 нм между поверхностями донора и акцептора аналогичное отношение может достигать величин ~ 0.1 . С другой стороны, полученные результаты свидетельствуют о том, что диполь-дипольное приближение вполне адекватно, когда переходы в доноре и акцепторе являются дипольно-разрешенными.

Литература

1. Агранович В.М., Галанин М.Д. Перенос энергии электронного возбуждения в конденсированных средах. – М.: Наука, 1978.

2. S. K. Lyo // *Phys. Rev. B* 62. – 2000. – 13641.
3. Š. Kos, M. Achermann, V. I. Klimov, and D. L. Smith // *Phys. Rev. B* 71. – 2005. – 205309.
4. C. R. Kagan, C. B. Murray, and M. G. Bawendi // *Phys. Rev. B* 54. – 1996. – 8633.
5. H. D. Robinson, B. B. Goldberg, and J. L. Merz // *Phys. Rev. B* 64. – 2001. – 075308.
6. S. A. Crooker, J. A. Hollingsworth, S. Tretiak, and V. I. Klimov // *Phys. Rev. Lett.* 89. – 2002. – 186802.
7. F. V. de Sales, S. W. da Silva, J. M. R. Cruz, A. F. G. Monte, M. A. G. Soler, P. C. Morais, M. J. da Silva, and A. A. Quivy // *Phys. Rev. B* 70. – 2004. – 235318.
8. A. Nazir, B. W. Lovett, S. D. Barrett, J. H. Reina, and G. A. D. Briggs // *Phys. Rev. B* 71. – 2005. – 045334.
9. G. D. Scholes and D. L. Andrews // *Phys. Rev. B* 72. – 2005. – 125331.
10. T. Pons, I. L. Medintz, M. Sykora, and H. Mattoussi // *Phys. Rev. B* 73. – 2006. – 245302.
11. G. Allan and C. Delerue // *Phys. Rev. B* 75. – 2007. – 195311.
12. D. M. Willard and A. van Orden // *Nature Materials*. – 2003. – 2. – 575.
13. X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weis // *Science*. – 2005. – 307. – 538.
14. S. Noda // *Science*. – 2006. – 260. – 314.
15. B. W. Lovett, J. H. Reina, A. Nazir, and G. A. D. Briggs // *Phys. Rev. B* 68. – 2003. – 205319.
16. S. Sangu, K. Kobayashi, A. Shojiguchi, and M. Ohtsu // *Phys. Rev. B* 69. – 2006. – 115334.
17. J. M. Taylor, H. A. Engel, W. Dür, A. Yacoby, C. M. Marcus, P. Zoller, and M. D. Lukin // *Nature Physics*. – 2005. – 1. – 177.
18. Imre, G. Csaba, L. Ji, A. Orlov, G. H. Bernstein, and W. Porod // *Science*. – 2006. – 311. – 205.
19. Y. Masumoto and T. Takagahara, eds. *Semiconductor Quantum Dots*. – Springer, Germany, 2002.
20. T. Förster // *Ann. Phys.* – 1948. – 437. – 55.
21. Ландау Л.Д., Лифшиц Е.М. *Квантовая механика: нерелятивистская теория*. Т. 3 – М.: Наука, 1988.
22. L. Dexter // *J. Chem. Phys.* – 1953. – 21. – 836.
23. J. Danckwerts, K. J. Ahn, J. Förstner, and A. Knorr // *Phys. Rev. B* 73. – 2006. – 165318.
24. T. Franzl, D. S. Koktysh, T. A. Klar, A. L. Rogach, J. Feldmann, and N. Gaponik // *Appl. Phys. Lett.* – 2004. – 84. – 2904.
25. T. Unold, K. Mueller, C. Lienau, T. Elsaesser, and A. D. Wieck // *Phys. Rev. Lett.* – 2005. – 94. – 137404.
26. A. Franceschetti and A. Zunger // *Phys. Rev. Lett.* – 1997. – 78. – 915.
27. E. L. Ivchenko and G. E. Pikus, *Superlattices and Other. Heterostructures. Symmetry and Optical Phenomena*. – Springer, Berlin, 1997.
28. Ансельм А.И. *Введение в теорию полупроводников*. – 2-е изд., М., Наука, 1978.
29. Варшалович Д.А., Москалев А.Н., Херсонский В.К. *Квантовая теория углового момента*. – Л.: Наука, 1975.
30. D. J. Norris and M. G. Bawendi // *Phys. Rev. B* 53. – 1995. – 16338.
31. *Properties of SiO₂ and Si₃N₄ at 300K* [Электронный ресурс]. – Режим доступа: <http://www.siliconfareast.com/sio2si3n4.htm>.
32. K. Vahala // *IEEE J. Quantum Electron.* – 1988. – 24. – 523.
33. A. A. Salman, A. Tortschanoff, M. B. Mohamed, D. Tonti, F. van Mourik, and M. Cherqui // *Appl. Phys. Lett.* – 2007. – 90. – 093104.
34. F. Gindele, K. Hild, W. Langbein, and U. Woggon // *J. Lumin.* – 2000. – 87-89. – 381.
35. S. F. Wuister, C. de Mello Donego, and A. Meijerinka // *J. Chem. Phys.* – 2004. – 121. – 4310.
36. V. I. Klimov and D. W. McBranch // *Phys. Rev. Lett.* – 1998. – 80. – 4028.

ВЛИЯНИЕ Br НА ФОТО-ТЕРМО-ИНДУЦИРОВАННУЮ КРИСТАЛЛИЗАЦИЮ СТЕКЛА

К.Е. Лазарева, Е.Ю. Акишина

Научный руководитель – д.ф.-м.н., профессор Н.В. Никоноров

В работе исследовалось влияние концентрации Br на фото-термо-индуцированную кристаллизацию стекла. При введении брома наблюдается смещение полосы поглощения коллоидного серебра в длинноволновую область спектра. Сделано предположение, что существует новый механизм образования кристаллической фазы, отличной от ранее известного.

Введение

Фото-термо-рефрактивные (ФТР) стекла – это сложный многокомпонентный материал, в состав которого входят фоточувствительные добавки (церий, серебро), активаторы, участвующие в «захвате» и «освобождении» фотоэлектронов (сурьма, олово) при температурах термопроявления, добавки, участвующие в формировании кристаллической фазы (фтор, бром), редкоземельные ионы, определяющие лазерные характеристики стекла (эрбий, иттербий) [1–5]. Они являются оптической средой, позволяющей записывать объемные фазовые голограммы, усиливать оптические сигналы и формировать на его основе планарные волноводные структуры. В процессе УФ-облучения и последующей термообработки в матрице стекла происходит изменение показателя преломления за счет выделения кристаллической наноразмерной фазы.

Следует отметить, что после длительного изучения этих стекол детальный механизм кристаллизации все еще неясен. В работах [6, 7] показано, что в образовании кристаллической фазы, содержащей NaF, существенную роль играет бром. В этих работах установлено, что в стеклах без Br нанокристаллы NaF не выделяются. В нашей работе изучена роль брома в кристаллизации ФТР-стекол, как УФ-облученных, так и необлученных.

Методика эксперимента

Стекла синтезировались в лабораторной силитовой печи при температуре 1450°C. Использовались материалы марки «хч» и «чда». Состав матрицы стекла: 17.7Na₂O–5.4ZnO–1.6Al₂O₃–75.2SiO₂–10.7Na₂O–0.01(Al₂O₃+ ZnO) (мол%). Остальные компоненты вводились сверх 100% в следующих количествах, мас. %: F – 2.5, Ag – 0.02, CeO₂ – 0.02, SnO₂ – 0.02, Sb₂O₃ – 0.1. На основе этой матрицы были синтезированы стекла, в которые вводились бромиды (в виде KBr) в количестве 0, 0.5, 1.5, 2 и 2.5 мас. % по синтезу. Получены образцы с толщинами от 0.69 до 0.73 мм.

Спектры оптического поглощения были измерены на спектрофотометре Cary 500 в интервале 200–1000 нм. Чтобы избежать влияния поверхностного загрязнения или поверхностной кристаллизации, каждый образец перед измерениями переполировывался. Измерение абсолютного показателя преломления каждого образца выполнялось на рефрактометре Abbe. Производились термообработки образцов при температурах 500°C и 520°C в течение 10 и 2 часов, соответственно.

Теоретическая часть

Ранее известный механизм предполагает, что изменение показателя преломления в ФТР стекле происходит вследствие фото-термо-индуцированной кристаллизации стекла, которая заключается в следующем (рис. 1).

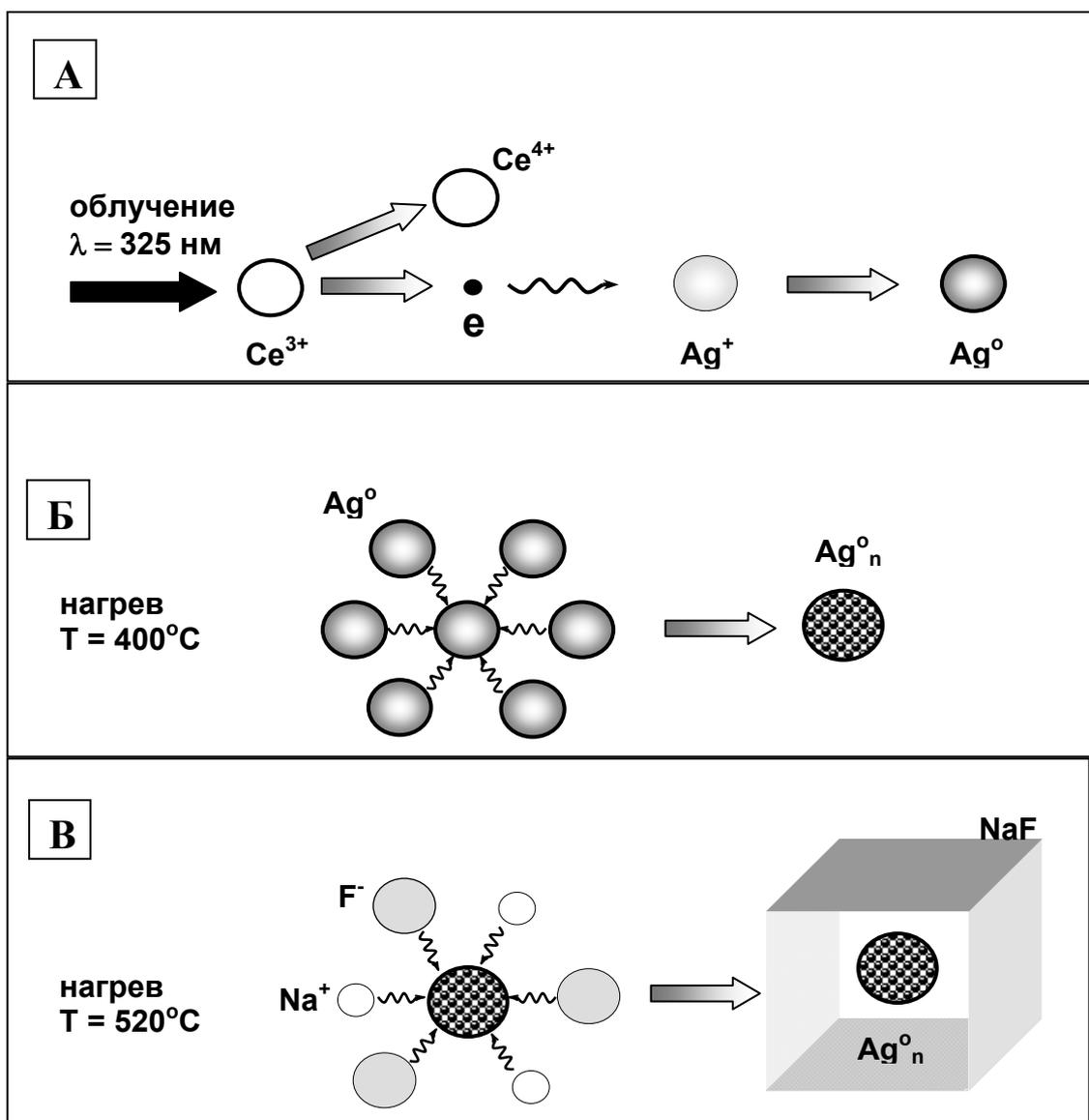
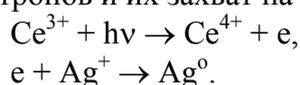


Рис. 1. Фото-термо-индуцированная кристаллизация стекла: А) фотоионизация церия УФ излучением с образованием свободных электронов, захват электронов ионами серебра и образование атомарного серебра; Б) образование коллоидного серебра при нагревании облученного стекла (400°C); В) рост микрокристаллов NaF на коллоидных центрах при 520°C

Под действием УФ-излучения с длиной волны 325 нм, близкой к полосе поглощения Ce^{3+} ($\lambda = 312 \text{ нм}$), происходит фотоионизация Ce^{3+} с образованием свободных электронов и их захват на ловушках Ag^+ (рис. 1А):



Последующая термообработка приводит к агрегации атомарного серебра Ag^0 и образованию коллоидных частиц серебра – Ag^0_n (рис. 1Б). При температурах, близких к температуре стеклования ($T_g = 490^\circ\text{C}$) и выше, на этих зародышах происходит рост микрокристаллов NaF (рис. 1В). Доза УФ-облучения определяет концентрацию коллоидных частиц серебра и, соответственно, концентрацию микрокристаллов, а температура и время термической обработки (термопроявление) определяют размер микрокристаллов и объемную долю кристаллической фазы галогенидов натрия.

В облученных стеклах показатель преломления (n_0) определяется двумя факторами: 1) показателем преломления кристаллической фазы (в данном случае для NaF

$n_{\text{кф}} = 1.32$) и ее объемной долей, 2) показателем преломления «остаточной» стеклофазы ($n_{\text{осф}} \approx 1.49$), из которой фтор и натрий перешли в кристаллическую фазу.

Можно отметить, что по отношению к показателю преломления (ПП) необлученной части стекла выделение нанокристаллов NaF ведет к уменьшению ПП, а рост нанокристаллов NaBr – к его увеличению. Поскольку концентрация F значительно превышает концентрацию Br и подвижность ионов фтора больше, чем подвижность ионов брома, то можно заключить, что при малых временах термообработки в первую очередь вырастают нанокристаллы NaF. Это ведет к уменьшению величины ПП в облученной части стекла (в максимуме интерференционной полосы) n_{max} и, соответственно, к увеличению δn . При длительных временах термообработки наряду с NaF могут вырастать микрокристаллы NaBr, что будет приводить к увеличению n_{max} и в итоге – к уменьшению δn [6].

Однако данный механизм образования нанокристаллов и их химический состав подвергается сомнению, поскольку результаты проведенных исследований противоречат этой теории. Возможен иной механизм образования кристаллической наноразмерной фазы. При термообработке образцов выше температуры плавления бромида серебра (434°C) происходит образование жидкого слоя AgBr на частицах Ag. Фторид натрия, NaF, растворяется в этом жидком слое, а при охлаждении из этого раствора происходит выделение кристаллов NaF.

Результаты и обсуждение

Влияние брома на спектр поглощения

На рис. 2 можно увидеть, что в УФ-области спектра наблюдается полоса поглощения Ce^{3+} , т.е. максимумы для всех концентраций брома находятся в интервале длин волн 300–305 нм. Установлено, что поглощение в УФ, видимой, ближней ИК-областях не показывает значительных изменений при различных концентрациях брома.

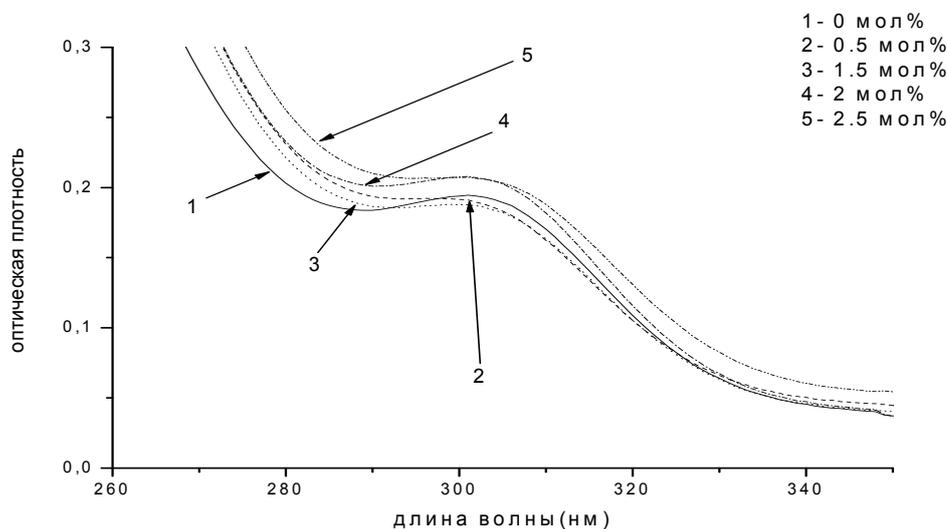


Рис. 2. Спектры поглощения исходных образцов с концентрациями 0–2.5 мол%

Далее мы изучали влияние брома на спектр поглощения облученных УФ образцов стекол в течение 15 минут, имеющих различную концентрацию брома (рис.3). Следует отметить, что УФ-экспозиция полностью перекрывает полосу поглощения Ce^{3+} . Для всех образцов спектры поглощения совпадают в пределах ошибки.

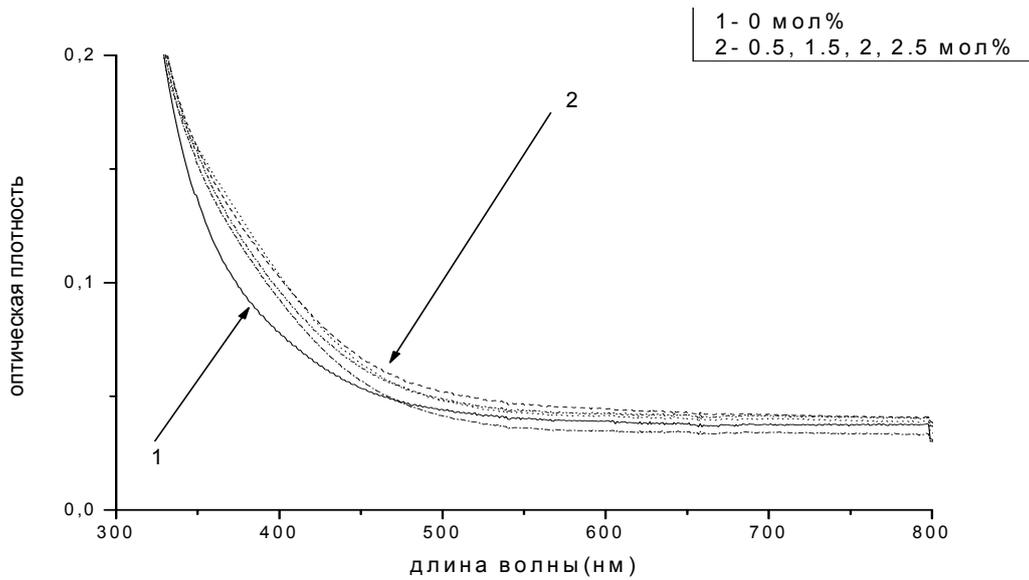


Рис. 3. Спектр поглощения облученных образцов

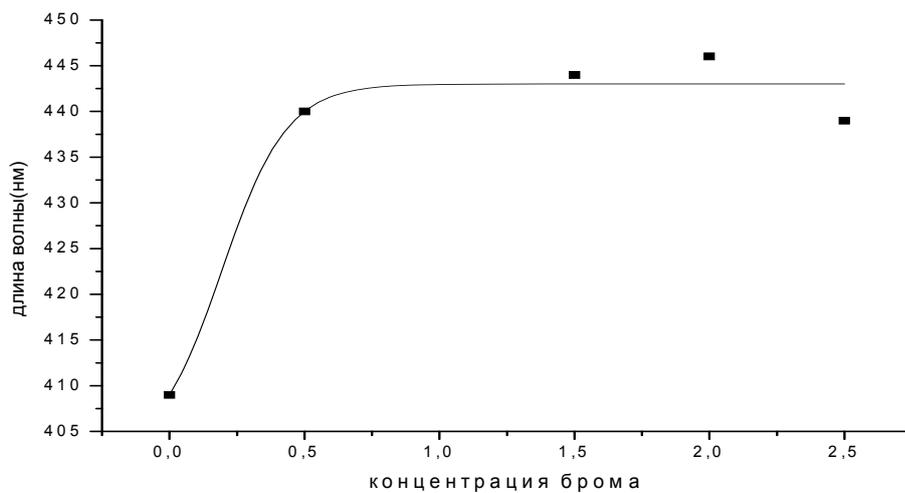
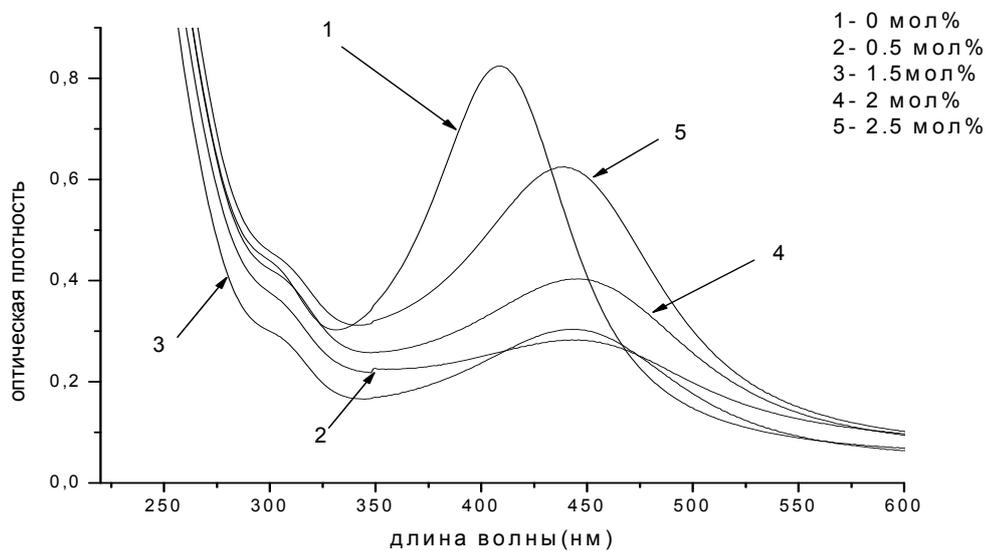


Рис. 4. а) Спектр поглощения УФ-облученных образцов после термической обработки при 500°C в течение 10 часов. б) Смещение максимума полосы поглощения в зависимости от концентрации брома

Важно отметить, что при термическом проявлении появляются атомы серебра, а кластеры, содержащие серебро, ответственны за дополнительные полосы поглощения, расположенные между 350 и 500 нм. На рис. 4, а, представлен спектр поглощения образцов после термической обработки при 500°. Можно увидеть, что полоса поглощения частиц, содержащих серебро, смещается в длинноволновую область с увеличением концентрации брома. Максимум полосы поглощения для образца без брома находится на 409 нм. С увеличением брома максимум смещается на 34 нм (рис.4, б). Смещение максимума полосы можно связать с двумя причинами: 1) изменение (увеличение) размера наночастиц серебра при термообработке стекла; 2) с образованием жидкого слоя вокруг нее, который состоит из бромида серебра. Однако конкретный вклад каждого механизма, в смещение полосы нуждается в дальнейших исследованиях.

При увеличении концентрации брома интенсивность полосы поглощения растет. Максимумы всех образцов очень размыты и, по-видимому, являются комбинацией двух различных полос: первая является следствием коллоидного серебра в матрице стекла, а вторая – серебра, покрытого слоем AgBr.

При последующей термообработке образцов при температуре 520° в течение 2 часов, максимумы наблюдаются на тех же длинах волн, но с большей интенсивностью.

Влияние брома на показатель преломления

Было изучено влияние брома на оптические свойства ФТР-стекол до и после облучения и термообработки. Измерения показателя преломления проводились на каждом этапе эксперимента.

На рис. 6 представлена зависимость показателя преломления от концентрации Br. Во всех образцах наблюдается общий характер увеличения показателя преломления при увеличении концентрации брома на всех стадиях обработки, что находится в хорошем согласии с результатами других работ [7].

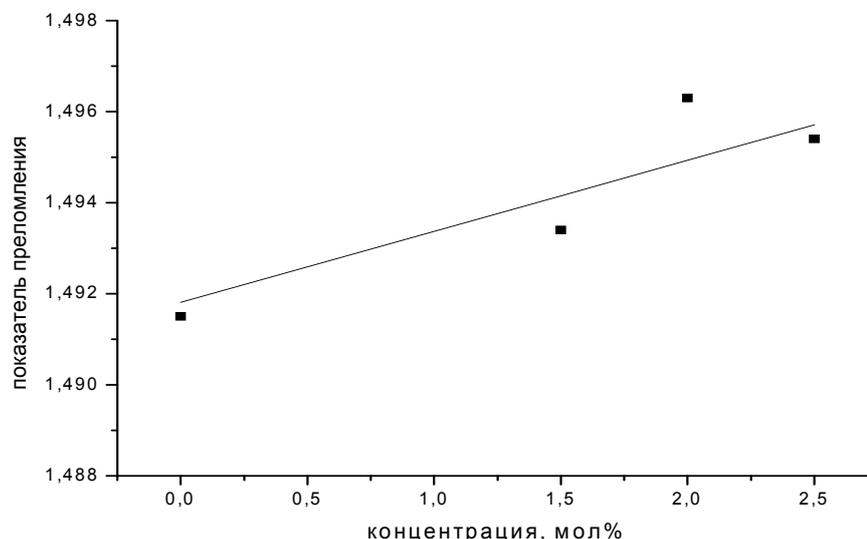


Рис. 6. Зависимость показателя преломления от концентрации брома для исходных, необлученных образцов

Выводы

Выяснено, что бром играет важную роль в создании серебро-содержащих ядер в УФ экспонированных стеклах. После термообработки при 500°C увеличение концентрации брома приводит к смещению полосы поглощения коллоидных частиц, серебра, в длинноволновую область спектра. При отсутствии брома наблюдается полоса по-

глощения на 409 нм, которая соответствует коллоидному серебру в матричном стекле. Установлено, что показатель преломления увеличивается с ростом концентрации брома, таким образом, высказано предположение, что серебряная частица окружена слоем AgBr, на которых выделяется кристаллический NaF.

Литература

1. Глебов Л.Б., Никоноров Н.В., Панышева Е.И., Туниманова И.В., Саввин В.В., Цехомский В.А. Фототерморефрактивное стекло // Труды VII Всес. Конф. по радиационной физике и химии неорганических материалов, ч. 2. – Рига: Изд. ИФ АН Латв. ССР, 1988. – С.527.
2. Никоноров Н.В., Панышева Е.И., Туниманова И.В., Саввин В.В. Мультихромные стекла – новая среда для оптической записи информации // Труды Всес. Конф. «Оптическое изображение и регистрирующие среды». – Л: Изд. ГОИ, 1990. – с.48.
3. Глебов Л.Б., Никоноров Н.В., Панышева Е.И., Петровский Г.Т., Саввин В.В., Туниманова И.В., Цехомский В.А. Мультихромные стекла – новые материалы для записи объемных фазовых голограмм // ДАН СССР. – 1990. – Т. 314. – № 4. – С. 849–853.
4. Кучинский С.А., Никоноров Н.В., Панышева Е.И., Туниманова И.В., Саввин В.В. Свойства объемных фазовых голограмм на мультихромных стеклах // Опт. и спектр. – 1991. – Т. 70. – № 6. – С.1296–1300.
5. Глебов Л.Б., Никоноров Н.В., Панышева Е.И., Петровский Г.Т., Саввин В.В., Туниманова И.В., Цехомский В.А. Новые возможности фоточувствительных стекол для записи объемных фазовых голограмм // Опт. и спектр. – 1992. – Т. 73. – № 2. – С.404–412.
6. Никоноров Н.В., Панышева Е.И., Туниманова И.В., Чухаев А.В. Влияние состава стекла на изменение показателя преломления при фототермоиндуцированной кристаллизации // Физика и химия стекла. – 2001. – Т 27. – № 3. – С. 365–375.
7. Glebova L., Lumeau J, Klimov M., Zanotto E., Glebov L.B. Role of bromine on the thermal and optical properties of photo-thermo-refractive glass // Journal of Non-Crystalline Solids. – 354 (2008). – P. 456–461.
8. Начаров А.П., Никоноров Н.В., Сидоров А.И., Цехомский В.А. Влияние УФ- и термообработки на морфологию наночастиц серебра в фото-термо-рефрактивных стеклах // Физика и химия стекла. – 2008 (в печати).

ОПТИЧЕСКИЕ СВОЙСТВА ФОТО-ТЕРМО-РЕФРАКТИВНЫХ СТЕКОЛ С РАЗЛИЧНЫМ СОДЕРЖАНИЕМ ФТОРА

Е.Ю. Акишина, К.Е. Лазарева

Научный руководитель – д.ф.-м.н., профессор Н.В. Никоноров

Синтезированы ФТР-стекла с разным содержанием фтора. Исследовано влияние фтора на спектры поглощения ФТР-стекла до и после УФ-облучения и термообработки. Показано, что введение фтора сдвигает полосу поглощения трехвалентного церия в коротковолновую область спектра. Установлено, что введение фтора в образцах после термообработки приводит к сдвигу полосы поглощения коллоидного серебра в длинноволновую сторону. Измерен показатель преломления ФТР-стекла. Показано, что с увеличением концентрации фтора показатель преломления ФТР-стекла уменьшается.

Введение

ФТР-стекла – это натрий-цинк-алюминий-силикатные стекла, содержащие церий, сурьму, фтор, серебро и бром, используемые для изготовления объемных дифракционных элементов для лазеров, оптической коммуникации и других оптических систем [1–2]. В этих стеклах после облучения УФ-радиацией и последующей термической обработки наблюдается изменение показателя преломления вследствие выделения кристаллической наноразмерной фазы фторида натрия.

Голограммы, зарегистрированные в ФТР-стекле, обладают высокой химической устойчивостью и механической прочностью и в этом отношении практически не отличаются от коммерческого оптического стекла К8. Кроме того, они выдерживают воздействие мощного непрерывного и импульсного лазерного излучения.

Существенным достоинством ФТР-стекла как материала для записи голограмм является его высокая однородность и воспроизводимость характеристик при синтезе самого стекла. ФТР-стекло, подобно оптическому стеклу К8, допускает применение традиционных методов механической обработки – шлифование и полирование, а также разнообразные технологии формования. Изготовление ФТР-стекла можно осуществлять как в лабораторных (до 600 г), так и в промышленных (до 300 кг) условиях с использованием стандартной оптической технологии.

Все эти достоинства еще раз подтверждают, что ФТР-стекла – многообещающий материал. Вместе с тем это достаточно новый оптический материал, имеющий много проблем для изучения и оптимизации. В данной работе исследуется, как влияет концентрация фтора, введенная в стекло при синтезе, на оптические и термические свойства ФТР-стекла.

Как отмечалось выше, в ФТР-стеклах после облучения УФ-радиацией и последующей термической обработки наблюдается изменение показателя преломления вследствие выделения кристаллической наноразмерной фазы фторида натрия. Заметим, что показатель преломления в облученной области меньше, чем в необлученной. Эта разница зависит от дозы облучения, температуры и времени термопроявления. Ранее было предложен следующий механизм образования кристаллов NaF [3]. Под действием УФ-излучения происходила фотоионизация Ce^{3+} с образованием свободных электронов и их захват на ловушках Ag^+ . Последующая термообработка приводила к агрегации атомарного серебра Ag^0 и образованию коллоидных частиц серебра – Ag_n^0 . При температурах близких к температуре стеклования ($T_g = 490^\circ\text{C}$) и выше, на этих зародышах происходит рост микрокристаллов NaF.

В данной статье предлагается несколько иной механизм образования кристаллической наноразмерной фазы фторида натрия. При термообработке образцов на коллоидных частицах серебра Ag_n происходит образование жидкого слоя. Состав этой жидкости мы пока не можем точно определить, но есть версии, что это AgBr или AgF , так как температуры плавления этих веществ незначительно отличаются друг от друга, а

именно 434°C и 435°C, соответственно. Фторид натрия, NaF, растворяется в этом жидком слое, а при охлаждении из этого раствора происходит выделение кристаллов NaF.

Экспериментальные результаты

Для проведения эксперимента была изготовлена серия образцов стекол, где концентрация фтора, изменялась следующим образом – 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол %. Образцы представляли собой прозрачные пластинки толщиной 0.74, 0.74, 0.73 и 0.75 мм соответственно. Далее исходные образцы подвергались УФ облучению в течение 15 минут и термической обработке 500°C – 10 часов и 520°C – 2 часа. После каждой обработки образцы полировались, чтобы избежать влияния поверхностных эффектов на результаты.

Спектры поглощения измерялись на всех стадиях эксперимента, и вычислялся условный коэффициент поглощения (оптическая плотность, нормированная к толщине 1 мм). Измерение спектров поглощения осуществлялось на спектрофотометре Cary500 в диапазоне длин волн 200-1000 нм с шагом 1 нм. Измерение показателя преломления на рефрактометре Аббе также проводилось на всех стадиях эксперимента.

На рис. 1 приведены спектры поглощения исходных образцов, где отчетливо выделяется пик полосы поглощения, соответствующий трехвалентному церию.

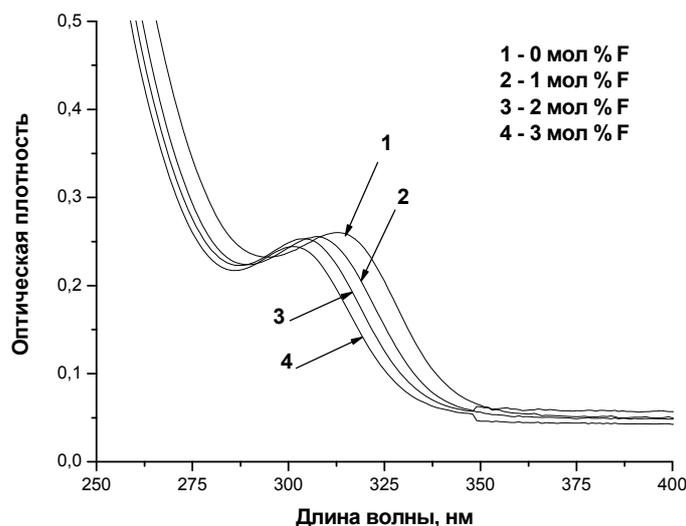


Рис. 1. Спектры поглощения исходных стекол с содержанием фтора 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол %

При увеличении концентрации фтора в стекле максимум этой полосы сдвигается в коротковолновую область спектра, а интенсивность полосы при этом незначительно уменьшается. Смещение полосы поглощения в коротковолновую часть спектра по мере увеличения концентрации введенного фтора можно объяснить появлением в окружении иона Ce^{3+} ионов фтора.

На рис. 2 приведена зависимость, которая показывает сдвиг максимума полосы поглощения трехвалентного церия в коротковолновую область при увеличении концентрации фтора.

Далее образцы облучались УФ-излучением в течение 15 минут. Спектр поглощения приведен на рис. 3.

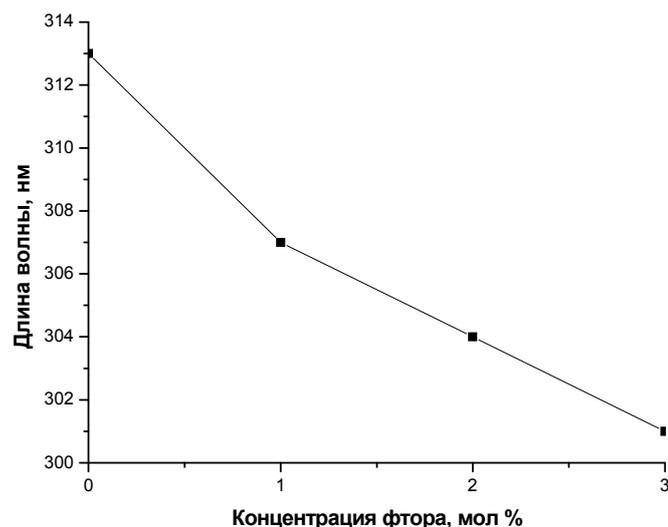


Рис. 2. Зависимость длины волны в максимуме полосы поглощения трехвалентного церия от концентрации фтора

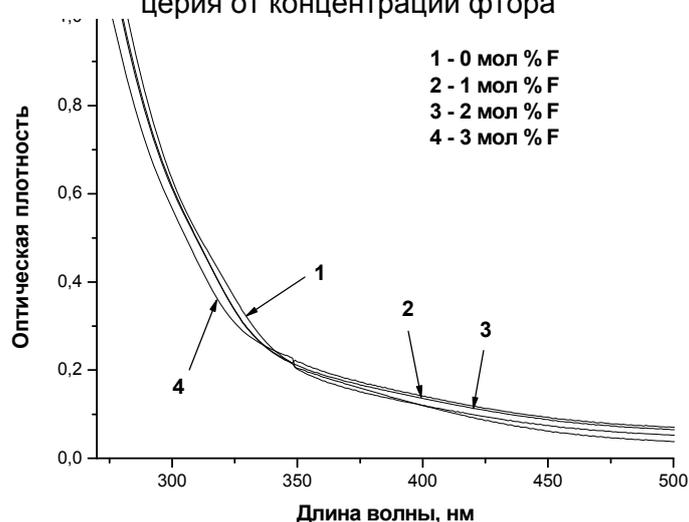
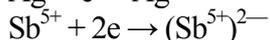
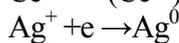
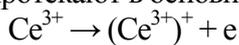


Рис. 3. Спектры поглощения стекол с содержанием фтора 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол % после облучения УФ

Характер изменения поглощения после УФ приведен на разностном спектре (рис. 4).

После УФ-облучения можно увидеть появление широкой полосы поглощения в области 250–300 нм и незначительное увеличение поглощения в области 350 нм. В процессе облучения протекают в основном следующие процессы:



Поглощение в УФ части спектра в основном связано с центрами $(\text{Ce}^{3+})^+$ и $(\text{Sb}^{5+})^{2-}$, тогда как поглощение в области 350–500 нм с квазимолекулами серебра типа Ag_2^+ [4].

Далее образцы термообрабатывались при температуре 500°C в течение 10 часов. В результате такой обработки возникла достаточно интенсивная окраска, связанная с выделением коллоидных частиц серебра. Облученная часть образца имеет спектр поглощения, представленный на рис. 5.

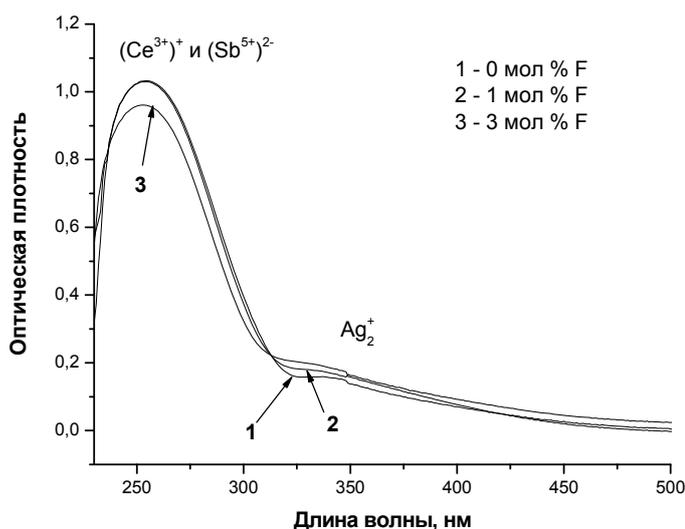


Рис. 4. Разностные спектры стекол после облучения УФ с содержанием фтора 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол %

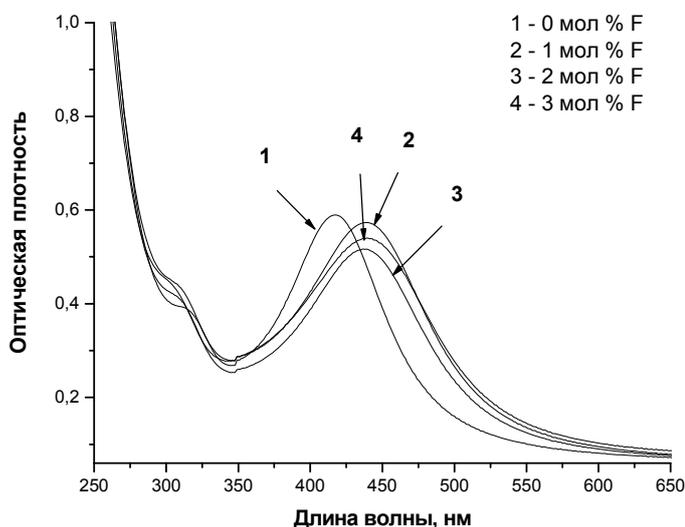


Рис. 5. Спектры поглощения стекол с содержанием фтора 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол % после термообработки 500°C в течение 10 часов облученной части образцов

На графике рис. 5 (кривая 1) максимум полосы поглощения образца без фтора лежит в области 418 нм. Известно, что в стекле без брома, но с фтором, после такой же термообработки (500°C в течение 10 часов) наблюдался пик поглощения на длине волны 409 нм [5]. Этот максимум был отнесен к коллоидному серебру в стекле. В нашем образце (рис. 5, кривая 1) нет фтора, но есть бром. Сдвиг в длинноволновую сторону на 9 нм по сравнению с [5] может быть связан со следующими причинами. В образце отсутствует фтор, который уменьшает показатель преломления, но присутствует бром, который его увеличивает. В итоге показатель преломления вокруг коллоидной частицы серебра растет, и полоса поглощения смещается в длинноволновую область.

В стеклах с содержанием фтора 1.0 мол %, 2.0 мол % и 3.0 мол % наблюдался сдвиг полосы поглощения в длинноволновую область (рис. 5). Наблюдаемый сдвиг в длинноволновую сторону полосы поглощения серебра в зависимости от концентрации фтора можно связать с несколькими механизмами.

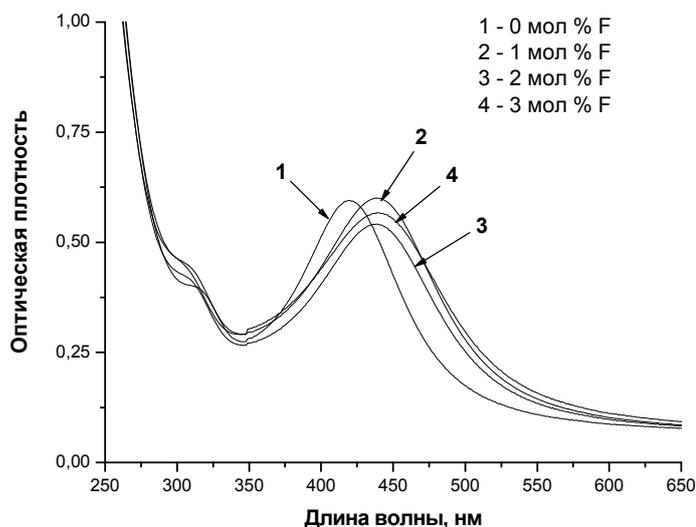


Рис. 6. Спектры поглощения стекол с содержанием фтора 0 мол %, 1.0 мол %, 2.0 мол % и 3.0 мол % после термообработки 520°C в течение 2 часов облученной части образцов

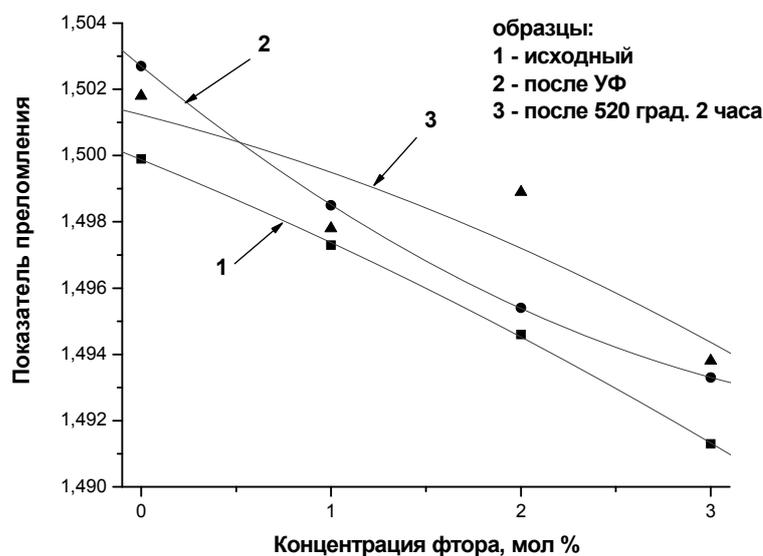


Рис. 7. Зависимость показателя преломления от концентрации фтора для образцов: исходного, облученного УФ и термообработанного при 520°C

1. Образуется слой, состоящий из AgBr и NaF, вокруг коллоидной частицы, который вносит вклад в увеличение показателя преломления. Но, как известно [3], при использовании ФТР-стекло для записи фазовых голограмм наблюдается эффект, связанный с отрицательным инкрементом показателя преломления. В то же время коэффициент термического расширения (КТР) области, состоящей из AgBr и NaF, образованной на коллоидном серебре, значительно выше, чем КТР матрицы стекла. Это означает, что окрестность этой области подвергается сильному напряжению, что вносит вклад в уменьшение показателя преломления. Можно ожидать, что такие напряжения могут создать вакуумные поры, которые будут вносить вклад в уменьшение показателя преломления [6].

2. При введении фтора вязкость стекла может уменьшаться. Соответственно будет увеличиваться подвижность всех ионов, включая ионы серебра. В результате размер

коллоидной частицы в стеклах, содержащих фтор, будет больше, чем без фтора. Увеличение размера коллоидной частицы серебра также может приводить к сдвигу полосы поглощения в длинноволновую сторону.

Высказанные механизмы нуждаются в дальнейших исследованиях.

Следующим этапом была термообработка при 520°C в течение 2 часов. Спектр поглощения приведен на рис. 6. Наблюдается незначительное увеличение интенсивности полос поглощения по сравнению с полосами после термообработки 500°C. В остальном характер спектра не изменился.

На каждой стадии обработки измерялся показатель преломления. На рис. 7 представлена зависимость показателя преломления от концентрации фтора. При увеличении концентрации фтора наблюдается уменьшение показателя преломления. Это согласуется с данными работы [7].

Заключение

Синтезированы ФТР-стекла с разным содержанием фтора. Исследовано влияние фтора на спектры поглощения ФТР-стекла до и после УФ-облучения и термообработки. Показано, что введение фтора сдвигает полосу поглощения трехвалентного церия в коротковолновую область спектра. Установлено, что введение фтора приводит к сдвигу полосы поглощения коллоидного серебра в длинноволновую сторону. Показано, что с увеличением концентрации фтора показатель преломления ФТР-стекла уменьшается.

Литература

1. Glebov L.B. Photosensitive glass for phase hologram recording // *Glastech. Ber. Glass Sci Technol.* – 71C (1998). – 85–90.
2. Efimov O.M., Glebov L.B., Glebova L.N., Richardson K.C., and Smirnov V.I. High-Efficiency Bragg Gratings in Photothermorefractive Glass // *Appl. Optics, Optical Technology and Biomedical Optics (OT&BO)*. – 1999. – 38. – 619–627.
3. Glebov L.B., Nikonorov N.V., Panysheva E.I., Petrovskii G.T., Savvin V.V., Tunimanova I.V., and Tsekhomskii V.A. New ways to use photosensitive glasses for recording volume phase holograms // *Opt. Spektrosk.* – 72. – August 1992. – 404–412.
4. Начаров А.П., Никоноров Н.В., Сидоров А.И., Цехомский В.А. Влияние УФ- и термообработки на морфологию наночастиц серебра в фототерморепрактивных стеклах // *Физика и химия стекла.* – 2008 (в печати).
5. Лазарева К.Е., Акишина Е.Ю., Цехомский В.А. Влияние брома на фото-терморепрактивный механизм стекол. // *Труды V Всероссийской межвузовской конференции молодых ученых* (в печати).
6. Glebov L.B. and Glebova L. Swelling of Photo-Thermo-Refractive Glass Resulted from Thermal Development // *Glastech. Ber. Glass Sci. Technol. C2* – 2002. – 75.
7. Китайгородский И.И. *Технология стекла.* – М., 1961. – 610 с.

ПОЛЯРИЗОВАННАЯ ЛЮМИНЕСЦЕНЦИЯ В СТЕКЛАХ И НАНОСТЕКЛОКЕРАМИКЕ, АКТИВИРОВАННЫХ ИОНАМИ ПЕРЕХОДНОГО ЧЕТЫРЕХВАЛЕНТНОГО ХРОМА

В.М. Ситдииков, С.Н. Жуков
Научный руководитель – В.А. Асеев

В работе была разработана и создана установка для регистрации поляризованной люминесценции, обнаружена поляризованная люминесценция четырехвалентного хрома в стеклах и наностеклокерамике. Также были измерены спектры поглощения исследуемых образцов.

1. Введение

В настоящее время большое внимание уделяется разработке сверхширокополосных (с шириной полосы более 100 нм) волоконных оптических усилителей и перестраиваемых лазеров для ближнего ИК диапазона (0.8–1.5 мкм). В этих целях перспективным является использование в качестве ионов-активаторов ионов переходных элементов, например хрома, имеющих широкие полосы вынужденного излучения (рис. 1) в ближней ИК области света (100–300 нм) [1–2]. Кроме того, сечения вынужденного излучения для ионов переходных металлов в несколько раз превышают сечения для редкоземельных ионов, используемых в настоящее время в волоконных усилителях. В связи с этим важную прикладную и фундаментальную проблему представляет определение валентных и координационных состояний хрома в различных матрицах.

Поэтому целью данной работы было экспериментальное исследование поляризованной люминесценции ионов хрома в стеклах и наноразмерных стеклокерамиках, которая позволяет определить валентное и координационное состояние переходных элементов в матрице. В работе была разработана и создана установка для регистрации поляризованной люминесценции, обнаружена поляризованная люминесценция четырехвалентного хрома в стеклах и наностеклокерамике. Показано, что степень поляризации значительно различается по всему диапазону регистрации и может служить дискриминатором валентного состояния переходных элементов в матрице.

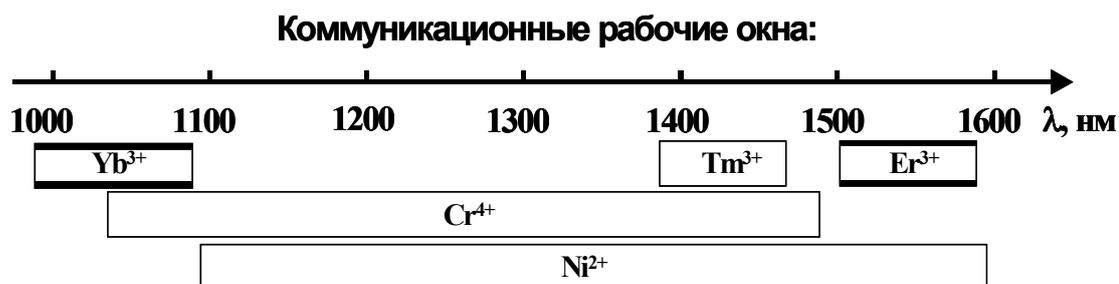


Рис. 1. Рабочие зоны различных ионов – активаторов

2. Методика эксперимента

2.1 Схема измерения спектров поглощения

Спектры поглощения для исследуемых образцов измерялись на спектрофотометре Varian Cary-500, схематично это представлено на рис. 2. Диапазон измерения 400–1300 нм. Использовался шаг измерения 1 нм и время интегрирования для одного измерения 0.1 с.

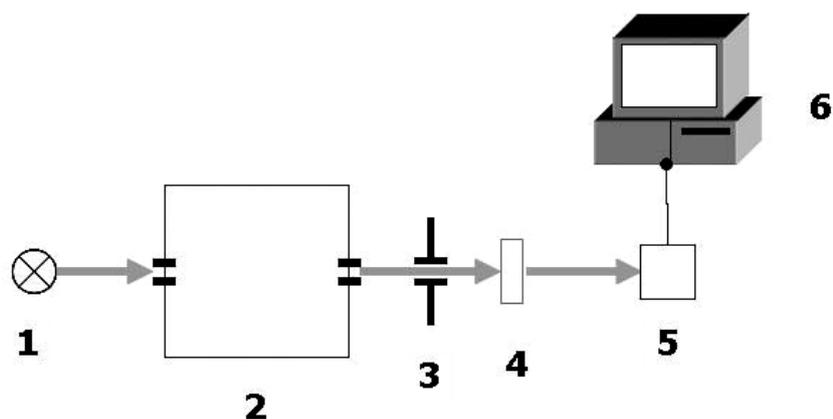


Рис. 2. Схема установки для измерения спектров поглощения: (1) – лампа, (2) – монохроматор, (3) – диафрагма, (4) – образец, (5) – приемник, (6) – компьютер

2.2 Поляризованная люминесценция

Анизотропия излучения и поглощения отдельного центра проявляет себя в эффекте поляризованной люминесценции. Поляризованная люминесценция состоит в анизотропии распределения электромагнитных колебаний в плоскости, перпендикулярной сигналу (рис. 3). Впервые поляризованная люминесценция исследовалась в кристаллах и органических соединениях П.П. Феофиловым [3].

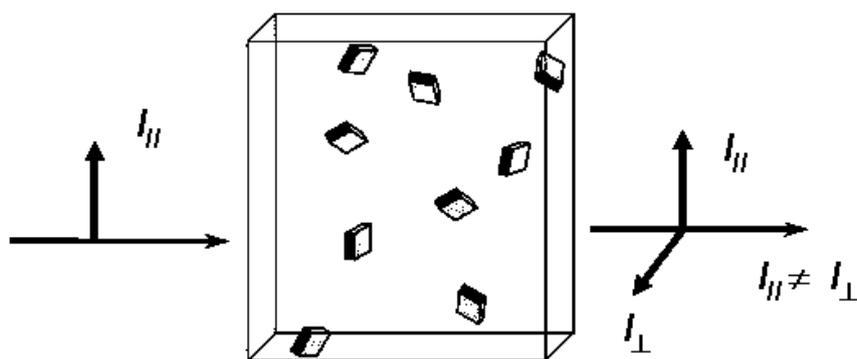


Рис. 3. Анизотропия излучения

В качестве основного экспериментального подхода применялось исследование поляризованной люминесценции, которая возникает при возбуждении линейно поляризованным светом. Так как оптические центры переходных элементов в неупорядоченных материалах не обладают симметрией, они испускают и поглощают свет анизотропно. Характер этой анизотропии существенно зависит от четности числа оптических электронов (электронов на d-оболочке для переходных элементов). Ввиду отсутствия симметрии вырождение энергетических уровней снимается максимально возможным образом. В случае четного числа оптических электронов вырождение снимается полностью, анизотропия поглощения или испускания света между любой парой уровней максимальна и соответствует линейному диполю. Такие ионы принято называть «некрамерсовыми». В случае нечетного числа оптических электронов у несимметричного оптического центра уровни дважды вырождены (образуют так называемые «крамерсовы дуплеты»), анизотропия поглощения или испускания света описывается в общем случае трехосным эллипсоидом и может, в частности, отсутствовать. Ионы с нечетным числом оптических электронов принято называть «крамерсовыми». Поэтому эффекты поляризованной люминесценции для некрамерсовых ионов гораздо больше, чем для крамерсовых. Благодаря указанному обстоятельству исследование поляризованной

люминесценции может послужить основой для селекции валентного состояния ионов d-элементов, в частности Cr^{3+} и Cr^{4+} .

Модель линейных осцилляторов дает хорошо известное объяснение эффекта поляризованной люминесценции, наблюдаемой при возбуждении линейно поляризованным светом [3]. Под действием такого света преимущественно возбуждаются оптические центры, у которых направление осциллятора, ответственного за поглощение, близко к направлению поляризации возбуждающего света. По отношению к возбужденным центрам образец приобретает аксиальную симметрию, и анизотропия его люминесценции соответствует такой симметрии. Таким образом, величина степени поляризации люминесценции определяется углом между поглощающим и излучающим осцилляторами.

2.3. Объект исследования и методика эксперимента

В качестве объекта исследования было выбрано синтезированное силикатное стекло ($\text{SiO}_2\text{-Al}_2\text{O}_3\text{-MgO-K}_2\text{O}$) с содержанием Cr_2O_3 0.4 вес.%. Для приготовления форстеритовой наностеклокерамики исходное стекло подвергалось термической обработке. При температурах 700°C (1–16 часов) и 800–900°C (1–12 часов) образовывалась нанокристаллическая фаза (Mg_2SiO_4) размером ~15–20 нм.

Следует специально отметить, что хром вводился в исходные материалы при синтезе в виде трехвалентного иона через оксид Cr_2O_3 . Но дальнейшая термическая обработка приводила к его окислению, и он переходил в четырехвалентное состояние. Об этом свидетельствуют литературные данные [4–6].

На рис. 4 изображена схема экспериментальной установки, на которой производились измерения спектров поляризованной люминесценции. Люминесценция возбуждалась в диапазоне 975–1010 нм излучением непрерывного титан-сапфирового лазера (модель 3900S, Spectra Physics) (2), который накачивался неодимовым лазером с удвоением частоты $\lambda=532$ нм (модель Millennia-Xs, Spectra Physics) (1). Также люминесценция возбуждалась и на 532 нм. Для изменения линейной поляризации возбуждающего света, излучение лазера пропускалось сначала через пластинку $\lambda/4$ (4), а затем через призму Глана (6).

Наблюдение сигнала люминесценции образца (7) производилось в направлении, перпендикулярном к направлению возбуждающего света. Люминесценция фокусировалась линзой (10) на щели монохроматора (модель Acton-300, Acton Research Corporation) (12) и регистрировалась приемником на InGaAs (модель ID-441, Acton Research Corporation) (13).

Свет люминесценции пропускался через закрепленный пленочный поляризатор (8), что позволяло измерять отдельно как $I_{\parallel}(\lambda)$ – интенсивность света люминесценции, поляризованного параллельно поляризации возбуждающего света, так и $I_{\perp}(\lambda)$ – интенсивность света люминесценции, поляризованного перпендикулярно поляризации возбуждающего света.

Для улучшения отношения сигнала к шуму использовалась система синхронного детектирования, реализованная на основе управляемого синхронизируемого усилителя (модель SR850, Stanford Research Systems) (14). Источником опорного сигнала служил модулятор (3). Результирующий продетектированный сигнал обрабатывался компьютером и представлялся как функция длины волны света люминесценции – поляризованный спектр $I(\lambda)$. Для улучшения дискриминации света люминесценции и возбуждающего света после поляризатора располагался ИК-фильтр (9). Для компенсации помех, вносимых монохроматором, был установлен кристалл шпата (11). Установка была проградуирована по спектральной чувствительности с помощью ленточной лампы.

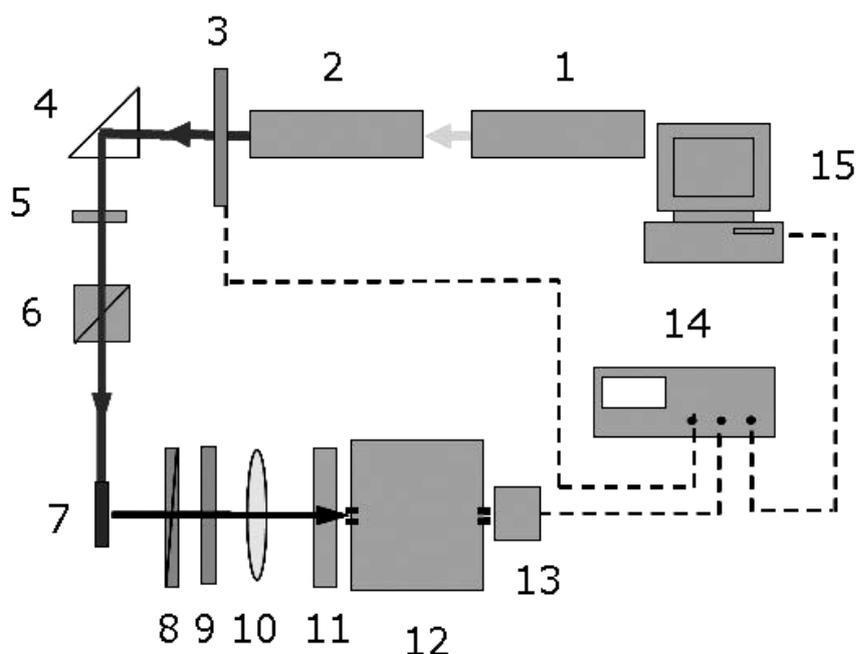


Рис. 4. Экспериментальная установка: (1) – Nd-лазер ($\lambda = 532$ нм), (2) – Ti-сапфировый лазер, (3) – модулятор, (4) – поворотная призма, (5) – пластинка $\lambda/4$, (6) – призма Глана, (7) – образец, (8) – пленочный поляризатор, (9) – ИК-фильтр, (10) – линза, (11) – кристалл шпата, (12) – монохроматор, (13) – InGaAs - приемник, (14) – синхронизированный усилитель, (15) – компьютер

3. Результаты и их обсуждение

По спектрам поглощения, которые показаны на рис. 5, видно, что после термической обработки появляется структура полос, характерных для иона Cr^{4+} в кристалле форстерита [7].

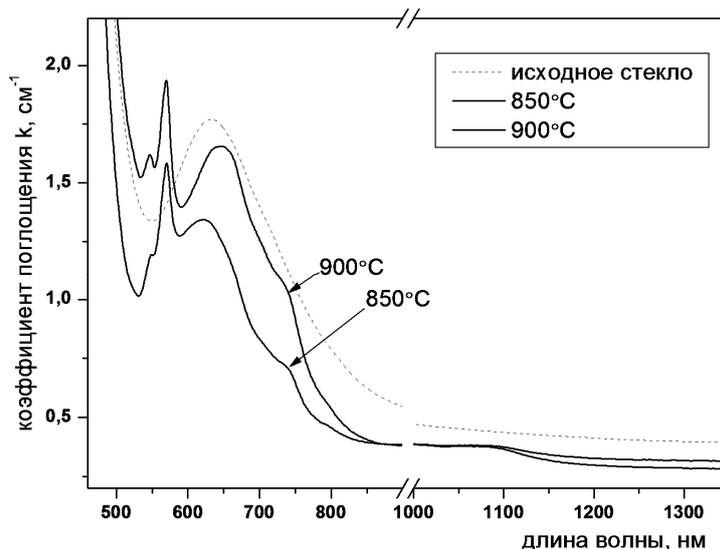


Рис. 5. Спектры поглощения исходного стекла и стеклокерамики

Измерения поляризованной люминесценции показали, что полоса люминесценции приобретает частичную поляризацию в интервале длин волн 900–1600 нм при возбуждении линейно поляризованным светом (рис. 6). Для стеклокерамики максимальные значения степени поляризации составляют $\sim 18\%$ и расположены в районе 1200–1300 нм.

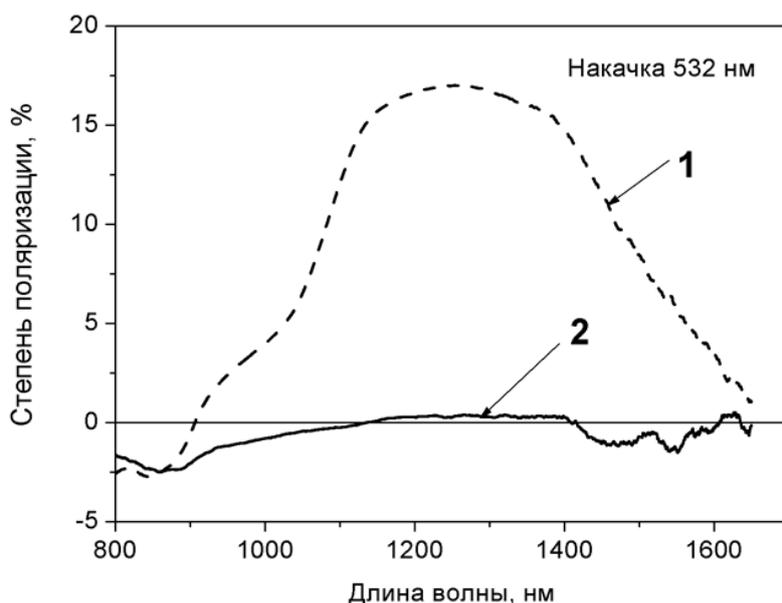


Рис. 6. Спектры поляризации люминесценции наностеклокерамики активированной хромом: (1) – спектр поляризации люминесценции, электрический вектор возбуждающего света перпендикулярен направлению наблюдения; (2) – спектр поляризации люминесценции, электрический вектор возбуждающего света параллелен направлению наблюдения

Как видно на рис. 7, контур люминесценции значительно отличается от контура поляризованной люминесценции. Этот эффект наблюдается как в стекле, так и в наностеклокерамике. Степень поляризации значительно разнится по всему диапазону регистрации. Поляризованная люминесценция наностеклокерамики намного меньше, когда электрический вектор возбуждающего света параллелен направлению наблюдения.

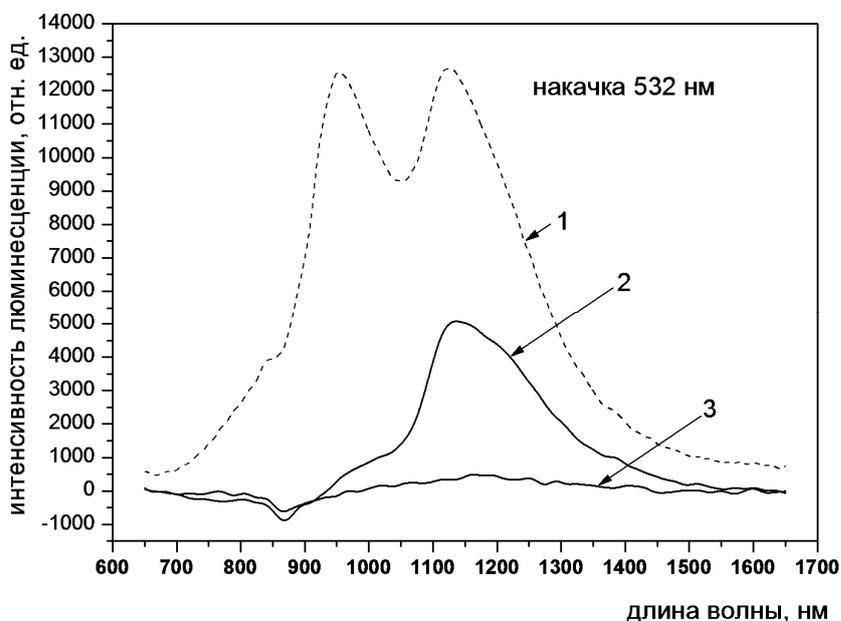


Рис. 7. Спектры люминесценции наностеклокерамики активированной хромом: (1) – спектр люминесценции; (2) – дифференциальный спектр поляризованной люминесценции, электрический вектор возбуждающего света перпендикулярен направлению наблюдения; (3) – дифференциальный спектр поляризованной люминесценции, электрический вектор возбуждающего света параллелен направлению наблюдения

Спектры люминесценции стекла, активированного хромом, показаны на рис. 8. Как видно, поляризация люминесценции у стекол меньше, чем у стеклокерамик, что согласуется с ранее описанной моделью линейных осцилляторов и служит подтверждением того, что центры четырехвалентного хрома, обладающие большей анизотропией по сравнению с трехвалентным, находятся в кристаллической фазе.

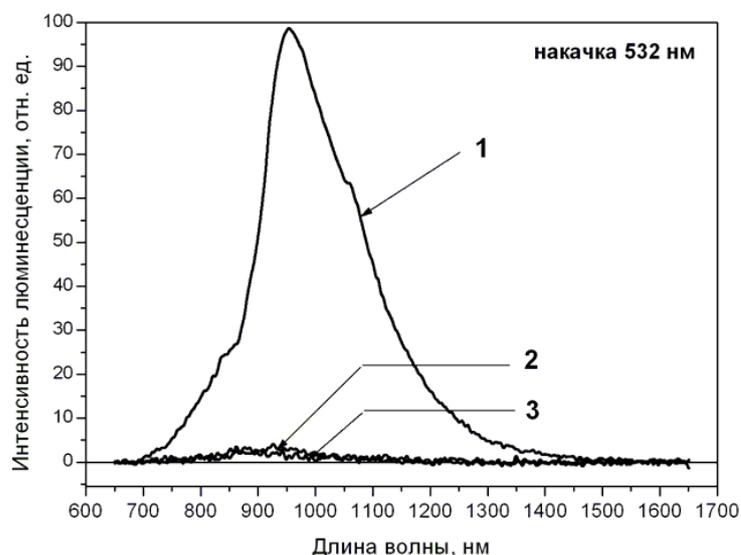


Рис. 8. Спектры люминесценции стекла, активированного хромом: (1) – спектр люминесценции; (2) – дифференциальный спектр поляризованной люминесценции, электрический вектор возбуждающего света перпендикулярен направлению наблюдения; (3) – дифференциальный спектр поляризованной люминесценции, электрический вектор возбуждающего света параллелен направлению наблюдения

4. Заключение

Полоса люминесценции ионов хрома в стеклах и форстеритовых наностеклокерамиках приобретает частичную поляризацию при возбуждении образцов линейно поляризованным светом, при этом степень поляризации люминесценции у стеклокерамик выше, чем у стекол. Степень поляризации значительно различается по всему диапазону регистрации. Форма контура дифференциального спектра значительно отличается от обычного спектра люминесценции. Все наблюдавшиеся эффекты могут быть удовлетворительно объяснены на основе модели, согласно которой при оптическом переходе между парой штарковских подуровней анизотропия поглощения и излучения света соответствует модели линейного осциллятора.

Литература

1. Felice V., Dussardier B. et al. Cr⁴⁺-doped silica optical fibers: absorption and fluorescence properties // *Eur. Phys. J. AP.* – 2000. – V.11. – P.107–110.
2. Beall G.N. Glass-ceramics for photonic applications // *Glastech. Ber. Glass Sci. Technol.* – 2000. – V.73. – C1. – P.3–11.
3. Феофилов П.П. Поляризованная люминесценция атомов, молекул и кристаллов. – М.: ГИФМЛ, 1959. – 288 с.
4. Pinckney L.R., Beall G.H. Transition element-doped crystals in glass // *Proc. SPIE.* – 2001. – V.4452. – P.93–99.

5. Petricevic V., Gayen S.K., Alfano R.R. Laser action in chromium-activated forsterite for near-infrared excitation: Is Cr^{4+} the lasing ion // Appl. Phys. Lett. – 1988. – V.53. – №26. – P. 2590–2592.
6. Verdun H.R., Thomas L.M. et al. Chromium-doped forsterite laser pumped with 1.06 μm radiation // Appl. Phys. Lett. – 1988. – V.53. – №26. – P.2593–2595.
7. Лебедев В.Ф., Гайстер А.В., Теняков С.Ю., Левченко А.Е., Дианов Е.М., Жариков Е.В., Спектрально-люминесцентные свойства сильнолегированных хромом монокристаллов форстерита. I. Спектры поглощения // Квант. электроника. – 2003. – 33 (3). – 192–196.

СПЕКТРАЛЬНО-ЛЮМИНЕСЦЕНТНЫЕ СВОЙСТВА ПРОЗРАЧНОЙ СВИНЦОВО-ФТОРИДНОЙ НАНОСТЕКЛОКЕРАМИКИ, АКТИВИРОВАННОЙ ИОНАМИ ЭРБИЯ

А.В. Клементьева

Научный руководитель – В.А. Асеев

Синтезирована прозрачная наностеклокерамика, активированная ионами эрбия, на основе свинцовофторосиликатного стекла. Показано, что термообработка свинцовофторосиликатного стекла приводит к образованию наноразмерной (16–40 нм) кристаллической фазы фторида свинца. На основании рентгенофазных исследований и измерения спектрально-люминесцентных свойств установлено, что ионы эрбия входят в кристаллическую фазу фторида свинца.

Введение

Прозрачные стеклокерамики представляют большой интерес для современной элементной базы фотоники. Занимая промежуточное положение между кристаллическими материалами и стеклами, они объединяют в себе лучшие свойства кристаллов (высокая механическая и термическая прочность) и стекол (возможность прессования и формования, возможность вытяжки оптического волокна и проведения ионного обмена для создания волноводных структур). Если активатор (например, эрбий, неодим и т.д.) входит в кристаллическую фазу, то спектрально-люминесцентные и лазерные характеристики стеклокерамики близки к характеристикам лазерных кристаллов-аналогов. Стеклокерамики – это гетерофазные структуры, которые формируются при отжиге стекла за счет роста кристаллической фазы в стеклообразной матрице. Одним из основных недостатков таких материалов является высокое светорассеяние на границе кристаллической фазы и стеклофазы. Поэтому ключевым направлением при разработке оптических стеклокристаллических материалов является уменьшение светорассеяния за счет роста наноразмерных кристаллов в матрице стекла.

Сегодня для создания лазеров, усилителей и конвертеров света представляют большой интерес прозрачные фторосодержащие наностеклокерамики, активированные редкоземельными ионами (Er, Eu, Tm, Nd, Pr) [1–10]. Редкоземельные ионы в окружении фторидов и, особенно, фторидов тяжелых металлов имеют меньшую энергию фонона. Такое бескислородное окружение редкоземельного иона уменьшает вероятность безызлучательной релаксации возбуждения. Это приводит к увеличению квантового выхода люминесценции по сравнению с кислородным окружением. Таким образом, прозрачные фторосодержащие наностеклокерамики, активированные редкоземельными ионами, являются перспективными средами для создания ап-конверсионных лазеров и широкополосных оптических усилителей. Так, например, в [11] показана возможность роста нанокристаллов фторида свинца (β -PbF₂), активированных эрбием, в матрице германо-оксифторидного и теллур-оксифторидного стекол, а также и исследованы спектрально-люминесцентные свойства таких нано-стеклокерамик в широком спектральном диапазоне (400–1600 нм). В [12] исследованы спектрально-люминесцентные свойства прозрачных свинцово-фторидных нано-стеклокерамик, активированных эрбием, полученных на основе оксидных свинцовофторосиликатных стекол, в видимом диапазоне. В работе приводятся результаты исследований спектров поглощения и люминесценции в видимом и ближнем ИК диапазонах для активированного эрбием исходного оксидного свинцовофторосиликатного стекла и PbF₃-нано-стеклокерамик, прошедших разные стадии термообработки.

Объект исследования и методика эксперимента

В работе синтезированы и исследованы образцы стеклокерамики свинцовофторо-силикатной системы $30\text{SiO}_2\text{-}18\text{PbF}_2\text{-}15\text{Al}_2\text{O}_3\text{-}5\text{ZnF}_2\text{-}29\text{CdF}_2\text{-}3\text{YF}_3$, активированной ErF_3 . Концентрация ErF_3 составляла 1 мол%, 1.5 мол% и 3.0 мол%. В таких системах при термообработке выделяется кристаллическая фаза, которая представляет собой фторид свинца $\beta\text{-PbF}_2$. В качестве центров нуклеации для кубической $\beta\text{-PbF}_2$ фазы выступают соединения ErF_3 , и, следовательно, ионы эрбия должны входить в кристаллическую фазу, образуя твердый раствор $\text{Pb}_{1-x}\text{Er}_x\text{F}_{2+x}$ [13].

Для синтеза стекол указанной системы применялись материалы марки «ОСЧ». Синтез проводился в течение 30 мин при температуре 1050°C в корундовых тиглях. При синтезе были предприняты специальные меры защиты от улета фторидных соединений. Синтез в корунде обеспечивал получение бесцветных прозрачных стекол с высоким уровнем пропускания в видимой области. Стекломасса вырабатывалась на холодную стеклоуглеродную форму и отжигались при температурах на 30°C ниже T_g (первичная термообработка).

Характеристические температуры были определены на основании данных дифференциального термического анализа (ДТА). Измерения ДТА проводили на дериватографе типа PAULIK-1500 фирмы "МОМ" (Венгрия). Результаты приведены на рис.1. Температура стеклования составила $T_g=430^\circ\text{C}$. Температура вторичной термообработки ($T=475^\circ\text{C}$) была выбрана в начале первого пика кристаллизации, который соответствовал выделению первой кристаллической фазы фторидов свинца.

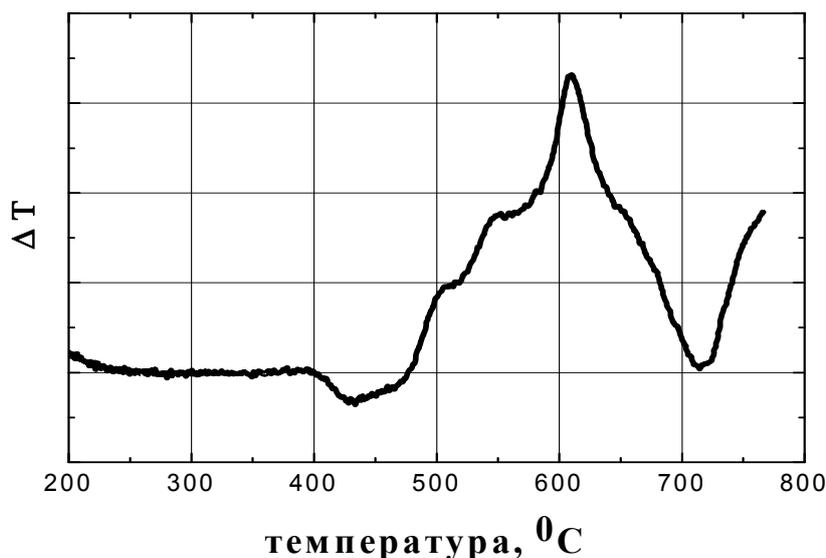


Рис. 1. Данные ДТА анализа

Образование нанокристаллов фторида свинца и переход ионов эрбия в кристаллическую фазу достигалась с помощью вторичной термообработки стекла. Главная задача такой обработки – создать относительно большую объемную долю наноразмерных кристаллов. Температура вторичной термообработки составляла 475°C , время 2–10 часов. Для определения состава и размеров кристаллической фазы проводился рентгенофазовый анализ образцов.

Спектры поглощения и люминесценции измерялись через каждый час термообработки. Спектры поглощения измерялись на спектрофотометре Varian Cary 500 в диапазоне длин волн 250–1700 нм с шагом 0,1 нм. Люминесценция образцов возбуждалась модулированным излучением неодимового лазера ($\lambda=532$ нм) (модель Millennia-Xs,

фирмы Spectra Physics). Далее излучение фокусировалось на входную щель монохроматора (модель Acton-300, фирмы Acton Research Corporation) и фотоприемник InGaAs (модель ID-441, фирмы Acton Research Corporation). Сигнал от фотоприемника усиливался и обрабатывался при помощи цифрового синхронного усилителя (модель SR850 фирмы Stanford Research Systems). Размеры образцов составляли $\sim 10 \times 10$ мм², толщина ~ 1 мм.

Для измерения кинетики затухания люминесценции использовано излучение импульсного лазера (модель LQ 129 фирмы Solar Laser system) на длине волны $\lambda_{\text{pump}} = 975$ нм. Кривые затухания люминесценции регистрировались цифровым запоминающим осциллографом (модель Infinium HP54830 фирмы Agilent Technologies). Время жизни определялось через отношение площади под кривой затухания к ее амплитуде.

Результаты и обсуждение

На первом этапе исследований была определена роль эрбия в формировании кристаллической фазы. Было установлено, что термообработка при температуре 475°C исходного стекла, не содержащего эрбий, не приводит к образованию кристаллической фазы. Одним из косвенных доказательств роста кристаллической фазы в стеклянной матрице является рассеяние света на границе раздела фаз. В спектрах ослабления таких гетерофазных материалов коротковолновая часть является наиболее чувствительной, и по ней можно судить о рассеивающих центрах – нанокристаллах, образующихся в результате термообработки. На рис. 2, а, показаны спектры ослабления стекла без эрбия до и после термической обработки в течение 10 часов. Эти спектры практически совпадают. На рис. 2, б, показан их разностный спектр. Изменения оптической плотности за счет рассеяния очень малы даже в коротковолновой части разностного спектра (менее 0.05). Отсутствие рассеяния в коротковолновой части спектра свидетельствует о том, что в стекле без эрбия кристаллическая фаза не выделяется даже при длительных временах термообработки.

Введение эрбия качественно меняет характер кристаллизации стекла. В этом случае ионы эрбия выполняют роль центров нуклеации. Рентгенофазовый анализ образцов после вторичной термообработки показал, что кристаллическая фаза представляет собой соединение состава $\text{Pb}_{1-x}\text{Er}_x\text{F}_{2+x}$. Рентгенофазовый анализ также показал, что для всех отмеченных концентраций ErF_3 размер элементарной ячейки кристаллов одинаков и составляет 5.75 Å. Таким образом, в результате вторичной термообработки стекла вырастает одна и та же кристаллическая фаза, которая не зависит от концентрации активатора.

Размер кристаллической фазы зависит от температуры и времени термообработки. На рис. 3 показана зависимость размера нанокристаллов от времени термообработки для концентрации $\text{ErF}_3 = 1.5$ мол%, построенная по данным рентгенофазового анализа. Увеличение времени термообработки ведет к увеличению размеров нанокристаллов. Так, при 2-часовой термообработке размер кристаллов составлял ~ 160 Å, а при 10-часовой ~ 400 Å. Дальнейшее увеличение времени термообработки не приводит к заметному увеличению размеров нанокристаллической фазы. Такие небольшие размеры нанокристаллов заметно снижают уровень светорассеяния в гетерофазной наностеклокерамике. Высокий уровень пропускания наностеклокерамики, а также однородность позволяет относить ее к оптическим материалам.

На рис. 4, а, показаны спектры оптических потерь, измеренные в диапазоне 350-1600 нм, для трех образцов: исходного стекла и двух наностеклокерамик, прошедших вторичную термообработку при 3 и 6 часах. Видно, что термообработка приводит к росту оптической плотности в коротковолновой области спектра и практически не изменяет ее в ИК диапазоне. Так, например, на длине волны 400 нм изменения оптиче-

ской плотности (разность между термообработанным и исходным образцами) составили $\Delta D = 0.08$ для времени термообработки $t=3$ часа и $\Delta D = 0.26$ для $t=6$ часа. В тоже время, для диапазона 1000-1600 нм $\Delta D \approx 0$. Небольшой рост оптической плотности в коротковолновой части спектра обусловлен светорассеянием на границе нанокристаллической фазы и стеклофазы.

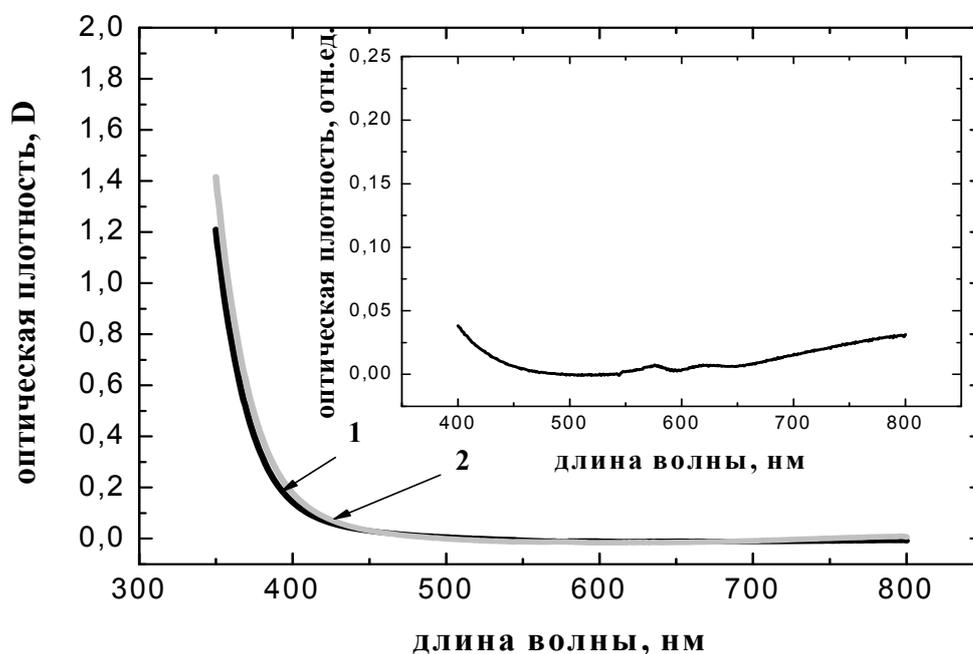


Рис. 2. а. Спектры оптического ослабления для исходного стекла (кривая 1) и стекла после термообработки в течении 10 часов при $T=475^{\circ}\text{C}$ (кривая 2).
б. Разностный спектр оптической плотности

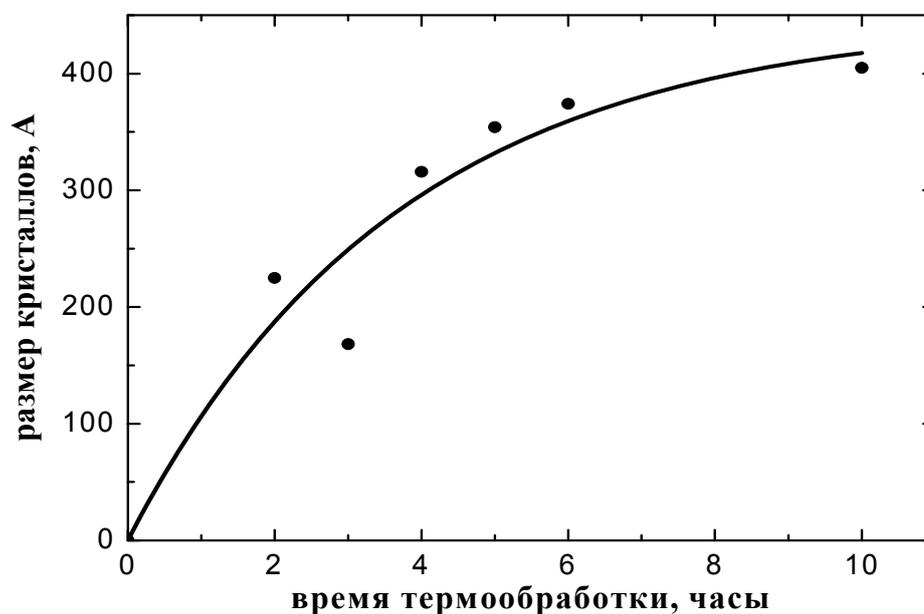


Рис. 3. Зависимость размеров кристаллов от времени температурной обработки при $T=475^{\circ}\text{C}$

Рассмотрим полосу поглощения, соответствующую основному лазерному переходу ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$ иона эрбия в диапазоне 1400–1700 нм. На рис. 4, б, показаны три конту-

ра поглощения – для исходного стекла и двух наностеклокерамик, для времени термообработки 3 и 6 часов. Все спектры нормированы по площади. При длительных временах термообработки (6 часов) в контуре наблюдается появление дополнительных штарковских полос и изменение соотношений их интенсивностей, характерное для кристаллической структуры. Такое поведение контура полос поглощения свидетельствует об изменении окружения эрбия при его переходе из стеклофазы в кристаллическую фазу.

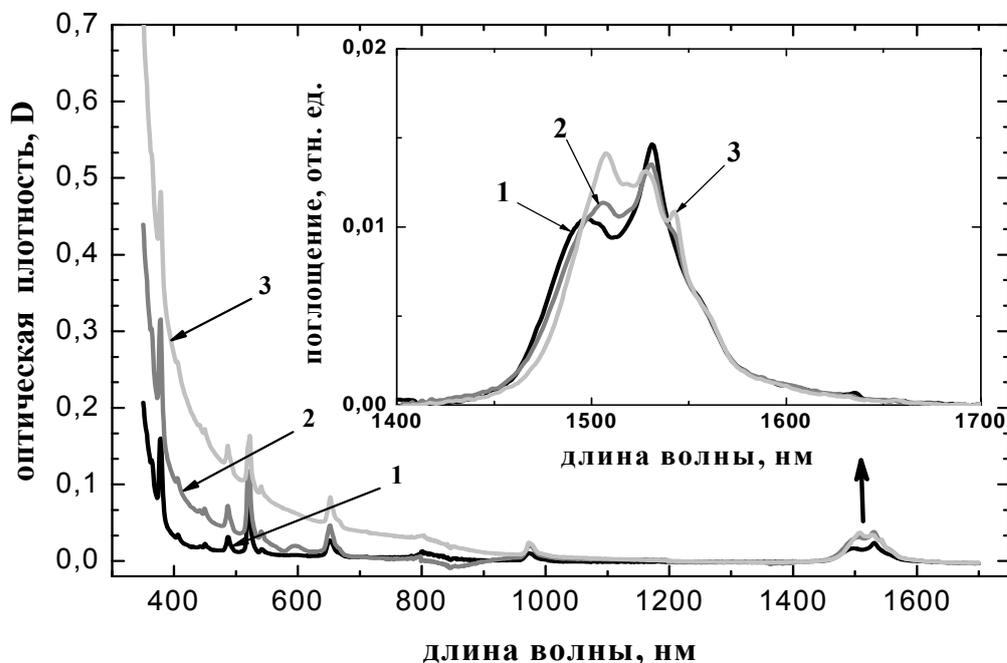


Рис. 4. а. Спектры ослабления исходного стекла, содержащего 1.5 мол% ErF_3 , (кривая 1) и наностеклокерамики, полученной после термообработки при $T=475^\circ\text{C}$ в течение 3 часов (кривая 2) и 6 часов (кривая 3). б. Фрагмент спектров поглощения для перехода ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$. Спектры нормированы по площади

На рис. 5, а показаны спектры ослабления образцов для трех концентраций ErF_3 , подвергнутых одинаковой термообработке в течение 6 часов при $T=475^\circ\text{C}$. Фрагменты спектров поглощения, соответствующие основному лазерному переходу ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$, показаны на рис. 5, б. Все спектры нормированы по площади. Обнаружено, что изменение концентрации фторида эрбия ведет к существенной деформации контура спектров поглощения. Такое необычное поведение контура полос поглощения от концентрации может объясняться изменением окружения эрбия при его переходе из стеклофазы в кристаллическую фазу.

В коротковолновой и видимой областях спектры для двух концентраций 1.0 и 1.5 мол% ErF_3 близки. Для этих двух концентраций в коротковолновой части наблюдается заметное увеличение оптической плотности, связанное с рассеянием на границе раздела кристаллической и стеклообразной фаз. Так, для длины волны 400 нм оптическая плотность для этих концентраций достигает $D=0.3$.

Спектр для концентрации 3.0 мол% существенно отличается от предыдущих двух. Рассеяние для этого образца мало. Так, для длины волны 400 нм оптическая плотность для концентрации 3.0 мол% на порядок величины меньше и составляет $D=0.03$.

Наблюдаемые изменения в спектрах при разных концентрациях фторида эрбия могут объясняться следующим образом. Ионы эрбия являются центрами кристаллизации. При малых концентрациях эрбия и постоянной концентрации «строительного материала» – фторидов свинца – вырастает небольшое количество нанокристаллов, при

этом они могут иметь большие размеры. Уровень светорассеяния в таких гетерофазных структурах высок (рис. 5, а, кривая 1). Увеличение концентрации ионов эрбия приводит к увеличению концентрации нанокристаллов и уменьшению их размеров в виду дефицита фторидов свинца. Уровень рассеяния будет определяться концентрацией и размером кристаллической фазы, т.е. он может снижаться или оставаться неизменным, как в случае с концентрацией 1.5 мол% (рис. 5, а, кривая 2). Дальнейшее увеличение концентрации ионов эрбия может приводить к случаю, когда «строительного материала» для всех ионов эрбия не хватает. В этом случае часть эрбия может находиться в кристаллической фазе или стеклофазе. При этом размеры кристаллической фазы малы, о чем свидетельствует низкий уровень светорассеяния в коротковолновой части спектра (рис. 5, а, кривая 3). В этом случае эрбий будет находиться в двух разных по составу и структуре окружения – кристаллическом и стеклообразном. Об этом свидетельствуют разные контуры поглощения на рис. 5, б.

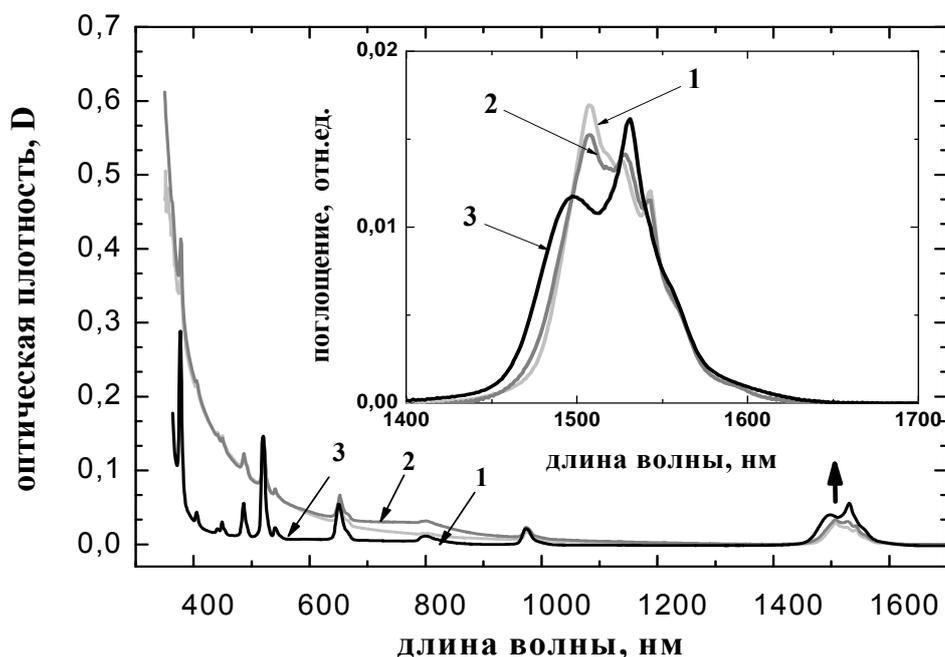


Рис. 5. а. Спектры ослабления трех наностеклокерамик с разным содержанием эрбия: ErF_3 – 1 мол% (кривая 1), 1.5 мол% (кривая 2), 3 мол% (кривая 3). Время термообработки $t=6$ часов. Температура обработки $T=475^\circ\text{C}$. б. Фрагмент спектров поглощения для перехода ${}^4I_{15/2} \rightarrow {}^4I_{13/2}$. Спектры нормированы по площади

На рис. 6 приведены нормированные по максимуму спектры люминесценции для исходного образца и двух наностеклокерамик, обработанных в течение 3 и 6 часов. Термообработка приводит к увеличению полуширины спектра. Так, полуширина спектра люминесценции ионов эрбия в исходном образце составляет 51 нм, в термообработанном – 69 нм. Уширение спектров может быть вызвано изменением окружения эрбия, т.е. часть ионов эрбия переходит в кристаллическую фазу и имеет окружение фторидов свинца, а остальная часть ионов эрбия остается в стеклообразной матрице, представляющей оксиды и фториды металлов. При больших временах термообработки (6 ч) доля эрбия, находящаяся в кристаллической фазе и в свинцовофторидном окружении увеличивается. Это приводит к проявлению штарковской структуры в спектрах люминесценции, характерной для кристаллов (рис. 6, кривая 3).

На рис. 7 показаны нормированные по максимуму спектры люминесценции образцов с разной концентрацией эрбия, подвергнутых одинаковой термообработке (3 часа). Образец с меньшей концентрацией (1 мол %) имеет наиболее выраженную штарк-

ковскую структуру в спектре люминесценции, характерную для спектров кристаллов. В спектре люминесценции для образца с высокой концентрацией (3 мол %) штарковская структура практически не просматривается, и спектр похож на спектр стекла.

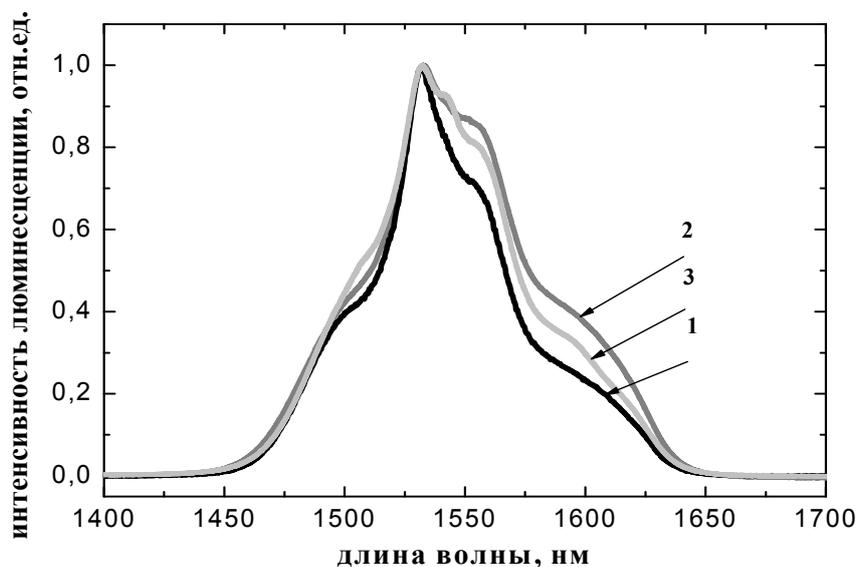


Рис. 6. Спектры люминесценции для исходного стекла, содержащего 1.5 мол% ErF_3 , (кривая 1) и наностеклокерамики, полученной при времени обработки $t=3$ часа (кривая 2) и $t=6$ часов (кривая 3). Температура обработки $T=475^\circ\text{C}$. Спектры нормированы по максимуму

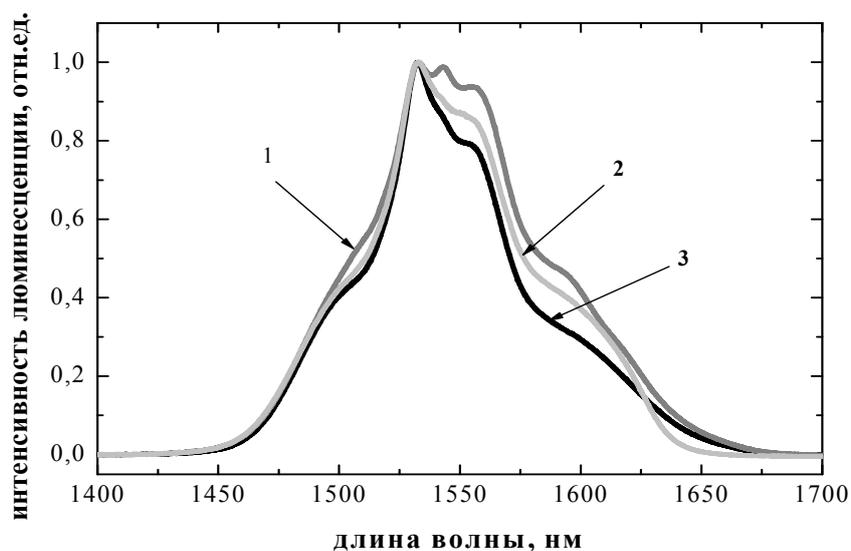


Рис. 7. Влияние концентрации ионов эрбия на спектр люминесценции термобработанного образца (время обработки – 3 часа)

Заключение

Синтезирована прозрачная наностеклокерамика, активированная ионами эрбия, на основе свинцовофторосиликатного стекла. Показано, что ионы эрбия выполняют роль центров нуклеации при росте свинцовофторидной кристаллической фазы. В процессе термообработки выделяется кристаллическая фаза, которая представляет собой соединение состава $\text{Pb}_{1-x}\text{Er}_x\text{F}_{2+x}$. По данным рентгенофазового и спектрально-люминесцентного анализа установлено, что эрбий входит в кристаллическую фазу.

Увеличение времени термообработки приводит к росту размера нанокристаллов, которые могут достигать 40 нм. Размер кристаллической фазы может уменьшиться с увеличением концентрации ионов эрбия. Обнаружено, что контур полосы поглощения для основного лазерного перехода может сильно искажаться в зависимости от концентрации эрбия, что может объясняться изменением окружения эрбия при его переходе из стеклофазы в кристаллическую фазу.

Литература

1. Y.H. Wang, J. Ohwaki // *Appl. Phys. Lett.* – 1993. – 63. – P.3268.
2. J. Mendez-Ramos, V. Lavin, I.R. Martin, U.R. Rodriguez-Mendoza, V.D. Rodriguez, A.D. Lozano-Gorrin, P. Nunez // *J.Appl. Phys.* – 2003. – 94. - P. 2295.
3. V.K. Tikhomirov, D. Furniss, A.B. Seddon, I.M. Reaney, M. Beggiora, M. Ferrari, M. Montagna, R. Rolli // *Appl. Phys. Lett.* – 2002. – 81. - P. 1937.
4. P.A. Tick, N.F. Borrelli, L.K. Cornelius, M.A. Newhouse // *J.Appl. Phys.* – 1995. – 78. – 6367.
5. Y. Kawamoto, R. Kanno, J. Qiu // *J. Mater. Sci.* – 1998. – 33. – P. 63.
6. M.A.P. Silva, V. Briois, M. Poulain, Y. Messaddeq, S.J.L. Ribeiro // *J. Phys. Chem. Sol.* – 2003. – 64. – №1. – P. 95.
7. M.J. Dejneka // *J. Non-Cryst. Solids.* – 1998. – 239. – №1–3. – P. 149.
8. H. Hayashi, S. Tanabe, T. Hanada // *J. Appl. Phys.* – 2001. – 89. – P. 1041.
9. M. Mortier, A. Monteville, G. Patriarche, G. Maze, F. Auzel // *Opt. Mater.* – 2001. – 16. – №1–2. – P. 255.
10. V.K. Tikhomirov, V.D. Rodriguez, J. Mendez-Ramos, P. Nunez, A.B. Seddon // *Opt. Mater.* – 2004. – 27. – P.543–547.
11. Z. Pan, A. Ueda, R. Mu and S.H. Morgan // *J. of Luminescence.* – 2007. – 126. – №1. – P. 251–256.
12. J. Mendez-Ramos, V. Lavin, I.R. Martin, U.R. Rodriguez-Mendoza, J.A. Gonzalez-Almeida, V.D. Rodriguez, A.D. Lozano-Gorrin, P. Nunez // *J. of Alloys and Compounds.* – 2001. – 323–324. – P.753–758.
13. G. Dantelle et al. // *J. of Solid State Chemistry.* – 2006 – 179. – P 2003–2011.

ВЛИЯНИЕ СУРЬМЫ НА ФОРМИРОВАНИЕ НАНОЧАСТИЦ СЕРЕБРА В ФОТО-ТЕРМО-РЕФРАКТИВНОМ СТЕКЛЕ

Е.В. Цыганкова

Научный руководитель – д.ф.-м.н., профессор Н.В. Никоноров

Синтезированы серебряные фототерморефрактивные стекла с переменным содержанием сурьмы от 0 до 0,45 мол.%. Исследована роль сурьмы в формировании коллоидных центров серебра в стекле. Показано, что коэффициент поглощения коллоидного серебра симбатно изменяется с содержанием сурьмы в стекле, а при отсутствии сурьмы в стекле коллоидное серебро в процессе термообработки не образуется.

Введение

Наноразмерные коллоидные частицы серебра, распределенные в матричном стекле, являются типичными плазмонами. Уникальные физические свойства плазмонов открывают огромные возможности их практического использования. Так, например, плазмоны считаются средством передачи информации на компьютерные чипы, так как они могут поддерживать очень высокую частоту. Их также рассматривают в качестве среды для высокоразрешающей литографии и микроскопии из-за их очень малой длины волны. Плазмоны также используются для усиления электромагнитного поля в рамановской спектроскопии. Кроме того, большое внимание уделяется поверхностным плазмонам в биотехнологиях [1].

При образовании коллоидного серебра изменяются спектральные и оптические свойства стекла – пропускание и показатель поглощения. Такой процесс можно использовать для записи объемных амплитудно-фазовых голограмм. В этом случае используют фото-термо-индуцированный процесс образования коллоидного серебра в матрице стекла [2] – под действием ионизирующего излучения в локальном месте стекла образуются центры коллоидообразования, и последующая термообработка приводит к росту коллоидных частиц серебра. Чтобы стекло было чувствительным к воздействию ионизирующего излучения, в него, как правило, вводят ионы церия, меди и другие ионы переменной валентности [2], являющиеся донорами электронов при воздействии излучения. Вместе с тем обязательным компонентом светочувствительных [2] и фототерморефрактивных стекол являются ионы сурьмы, роль которых недостаточно выяснена. Целью настоящей работы явилось исследование роли сурьмы в формировании наноразмерных серебряных коллоидных центров (плазмонов) в фоточувствительном стекле.

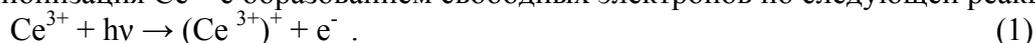
Методика эксперимента

Для проведения исследований в работе синтезированы стекла системы $\text{Na}_2\text{O} - \text{Al}_2\text{O}_3 - \text{ZnO} - \text{SiO}_2$ с переменным содержанием сурьмы от 0 до 0,45 мол.%. Стекла синтезировались в кварцевых тиглях при температуре 1480°C в воздушной атмосфере. Температура стеклования (T_g) составляла 495°C . Образцы толщиной 1 мм в виде полированных плоскопараллельных пластин облучались УФ-излучением и термически обрабатывались.

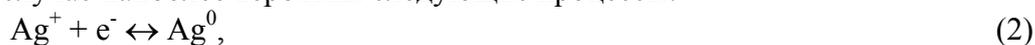
Облучение УФ осуществлялось ртутной лампой высокого давления. Исследования проводились для следующих доз облучения: 10, 100, 1000 и 10000 Дж. Из спектра излучения ртутной лампы интерференционными фильтрами вырезалась область 290–410 нм. Образец облучался через диафрагму диаметром 5 мм. Термообработка проводилась в муфельной печи при температуре $560 \pm 2^\circ\text{C}$ 3 часа. Спектры поглощения измерялись на всех стадиях эксперимента: 1) у исходного стекла, 2) после УФ-облучения исходного стекла, 3) после термической обработки исходного и облученного стекла. Измерение спектров поглощения осуществлялось на спектрофотометре Cary 500 в диапазоне длин волн 200–1000 нм с шагом 1 нм.

Результаты и обсуждение

Первым этапом работы явилось исследование спектров поглощения образцов стекол после УФ-облучения. Под действием УФ-излучения в данных стеклах происходит фотоионизация Ce^{3+} с образованием свободных электронов по следующей реакции [3]:



Освободившиеся электроны захватываются на ловушках, которыми могут являться Ag^+ , Sb^{5+} , Ce^{4+} , анионные вакансии стекла и другие дефекты и примеси стекла. В нашем случае наиболее вероятны следующие процессы:



Атомарное серебро Ag^0 термически нестабильно и при нагревании может разрушаться до Ag^+ [4]. В то же время заряженный центр $(Sb^{5+})^-$ может сохраняться при высоких температурах [5].

На рис. 1 приведены спектры поглощения исходных стекол с разным содержанием сурьмы. Наблюдаемые изменения в спектрах очень небольшие, однако, можно заметить определенные тенденции – переход от стекла, не содержащего сурьму, к стеклу, содержащему 0,048 мол.% Sb_2O_3 (по синтезу), приводит к смещению границы поглощения в коротковолновую часть спектра. Это можно связать с восстановлением примесей стекла (например, Fe_2O_3) при введении сурьмы. Дальнейшее увеличение содержания сурьмы приводит к смещению границы поглощения в длинноволновую часть спектра. Этот результат согласуется с работой [6].

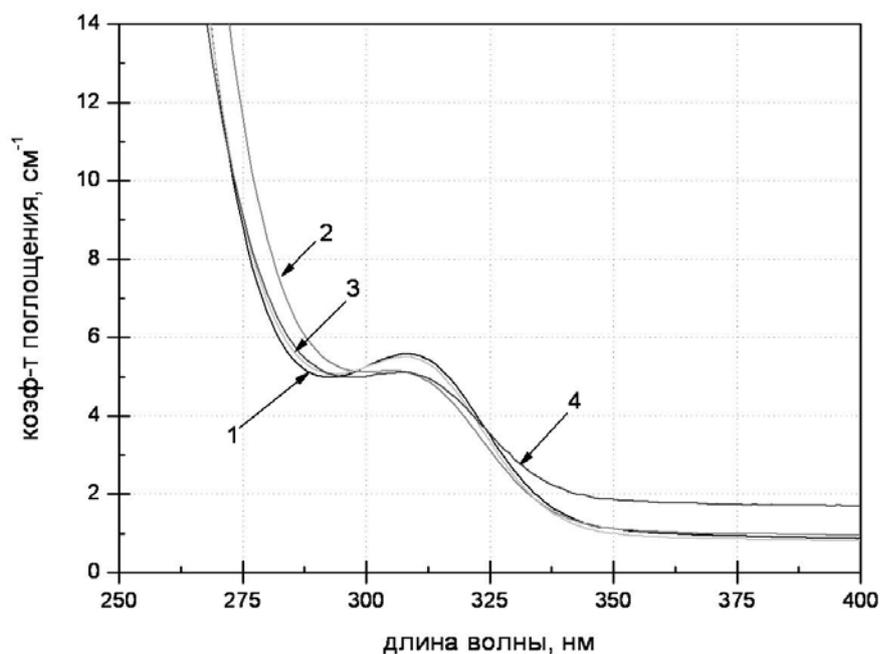


Рис. 1. Спектры поглощения исходных стекол, содержащих Sb_2O_3 (кривая 1 – 0 мол.%, 2 – 0,048 мол.%, 3 – 0,15 мол.%, 4 – 0,45 мол.%)

На рис. 2 приведены спектры поглощения, полученные при различных дозах облучения образца стекла, содержащего 0,15 мол.% сурьмы. Для стекол, содержащих 0, 0,048 и 0,45 мол.% сурьмы, наблюдаются похожие зависимости.

Из графиков видно, что при увеличении дозы облучения увеличивается поглощение в области 250–450 нм и сдвигается край фундаментального поглощения в длинноволновую область.

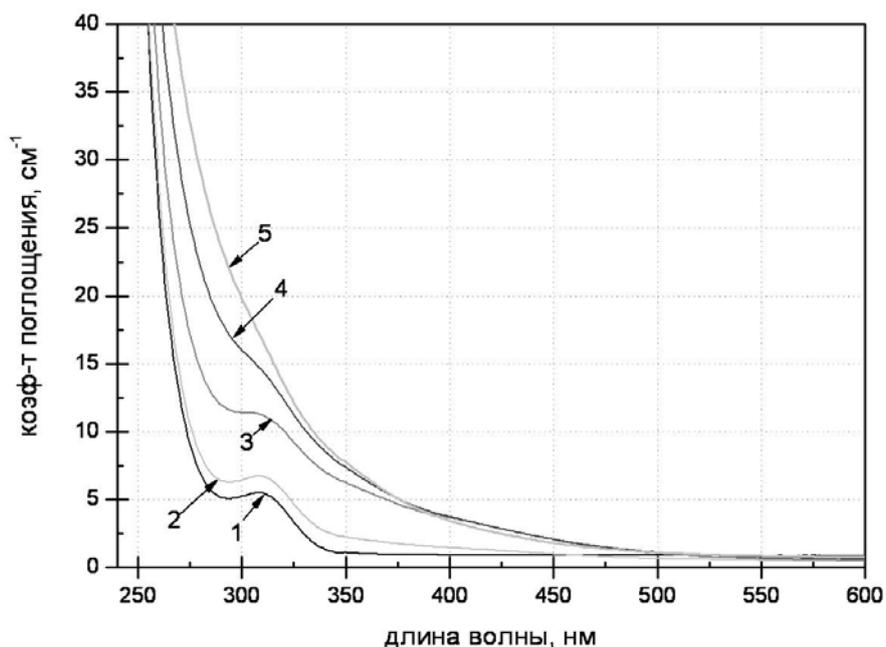


Рис. 2. Спектр поглощения стекла с содержанием сурьмы 0,15 мол.% при различных дозах облучения УФ (кривая 1 – необлученный образец, 2 – облученный образец 10 Дж, 3 – 100 Дж, 4 – 1000, 5 – 10000 Дж)

Следует отметить, что изменение содержания сурьмы в стекле не оказывает существенного влияния на спектры, наведенные при облучении. Во всех исследованных стеклах наблюдается рост поглощения и смещение края фундаментального поглощения. Учитывая, что при последующей термической обработке поведение этих стекол в корне изменяется, была предпринята попытка более детально изучить фотоиндуцированные спектры.

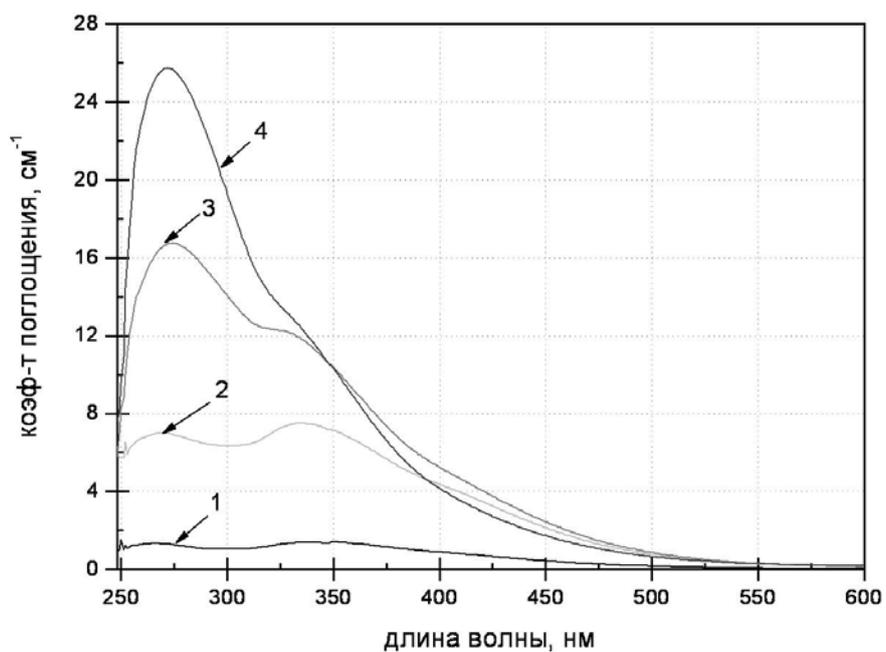


Рис. 3. Спектры добавочного поглощения стекла без сурьмы при различных дозах облучения (10 Дж – кривая 1, 100 Дж – кривая 2, 1000 Дж – кривая 3, 10000 Дж – кривая 4)

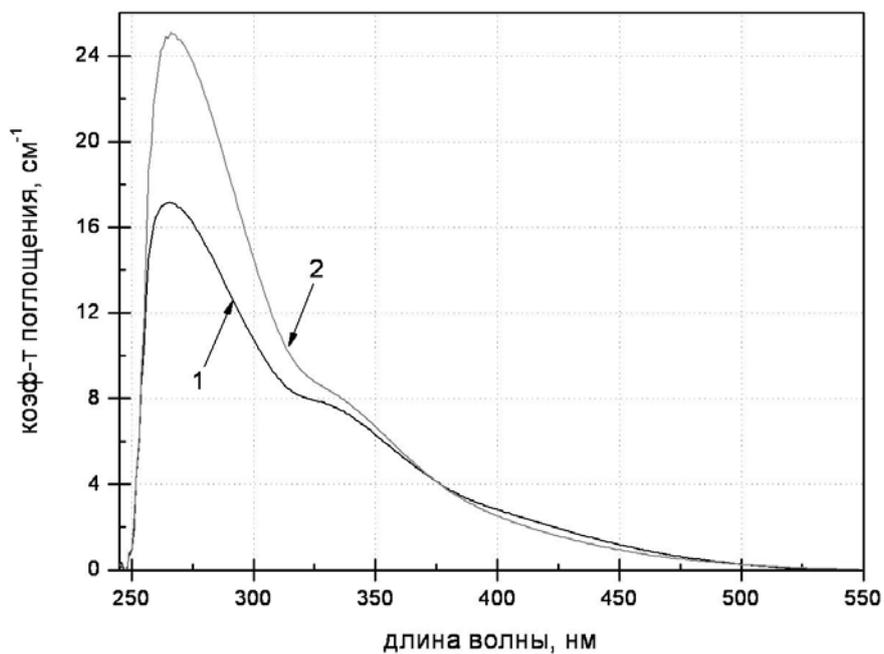


Рис. 4. Спектры добавочного поглощения стекол с содержанием сурьмы (кривая 1 – 0 мол.%, 2 – 0,15 мол.%), облученных дозой 1000 Дж

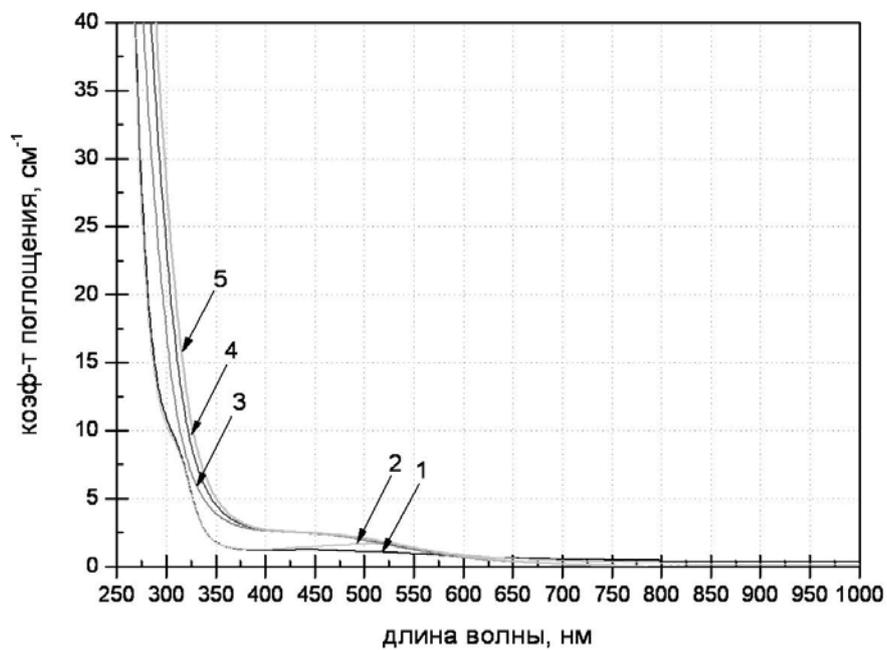


Рис. 5. Спектр поглощения стекла без сурьмы после термообработки 560°C (кривая 1 – необлученный образец, 2 – облученный образец 10 Дж, 3 – 100 Дж, 4 – 1000, 5 – 10000 Дж)

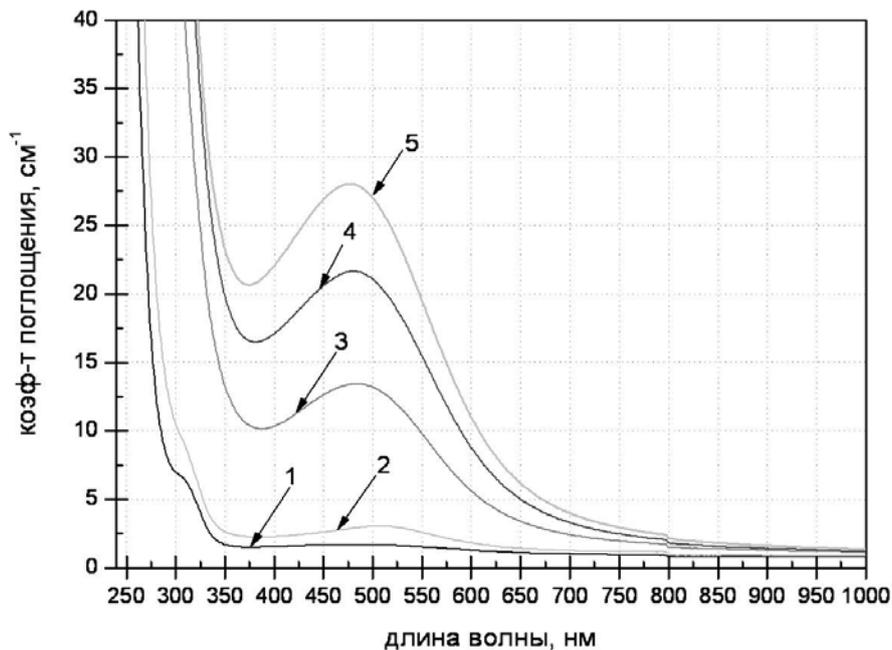


Рис. 6. Спектр поглощения стекла с содержанием сурьмы 0,048 мол.% после термообработки 560°С (кривая 1 – необлученный образец, 2 – облученный образец 10 Дж, 3 – 100 Дж, 4 – 1000, 5 – 10000 Дж)

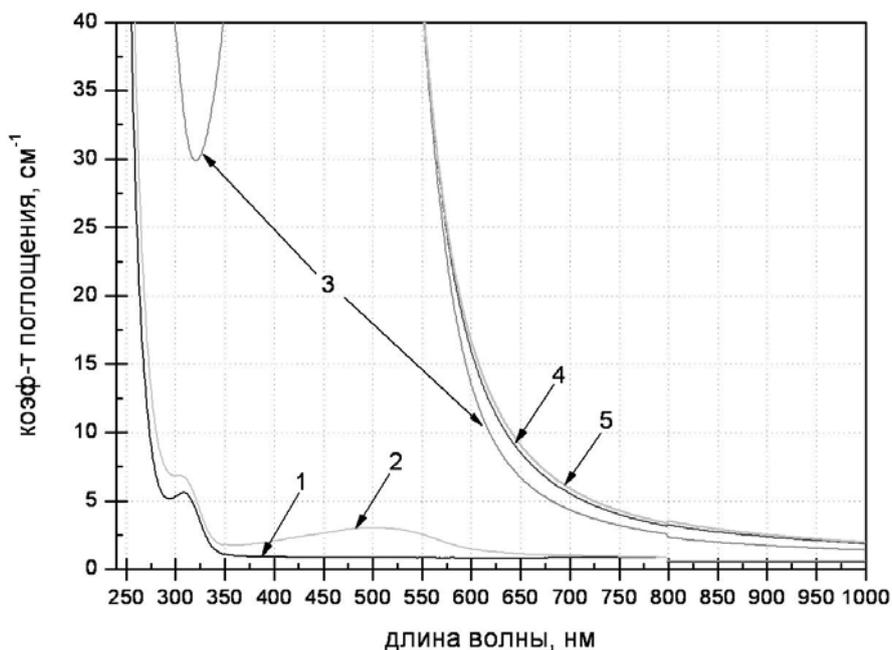


Рис. 7. Спектр коэффициента поглощения стекла с содержанием сурьмы 0,45 мол.% после термообработки 560°С (кривая 1 – необлученный образец, 2 – облученный образец 10 Дж, 3 – 100 Дж, 4 – 1000, 5 – 10000 Дж)

На рис. 3 приведены спектры добавочного поглощения для стекла без сурьмы при разных дозах облучения. Спектры добавочного поглощения представляют собой разность в спектрах между облученным и необлученным стеклом. Анализ этих спектров позволяет увидеть две широкие полосы поглощения в области 270 и 330 нм. С ростом дозы облучения интенсивность полос поглощения возрастает. Появление этих добавочных полос можно связывать с несколькими причинами: ростом полосы поглощения

(Ce^{3+})⁺ в области 270–280 нм (реакция 1) и ростом широкой полосы поглощения атомарного серебра с максимумом в районе 405 нм (реакция 2).

На рис. 4 показаны спектры добавочного поглощения для двух стекол – с сурьмой (0.15 мол.%) и без сурьмы, облученных одинаковой дозой 1000 Дж. Следует отметить, что интенсивность полосы в области 270 нм для сурьмяного стекла больше, чем для стекла без сурьмы. Дополнительный рост поглощения на 270 нм может быть связан с захватом фотоэлектрона на Sb^{5+} и образованием комплекса (Sb^{5+})⁻ по реакции (3). Эффект увеличения поглощения в районе 270 нм нами наблюдался для всех стекол, содержащих сурьму.

На рис. 5–7 приведены спектры поглощения стекол после облучения УФ и последующей термообработки при 560°C. Анализ графиков показывает, что при термообработке образуется коллоидное серебро с максимумом поглощения в районе 450 нм. Интенсивность этой полосы возрастает с увеличением дозы облучения. Следует также отметить, что коэффициент поглощения коллоидного серебра увеличивается с увеличением содержания сурьмы в стекле, а при отсутствии сурьмы коллоидное серебро в процессе термообработки не образуется. Т.е. полоса поглощения с максимумом 450 нм отсутствует в стекле, не содержащем сурьму (рис. 5).

Наблюдаемые эффекты объясняются следующим образом. При нагревании стекла без сурьмы атомарное серебро разрушается, т.е. идет процесс



Освободившийся термический электрон может захватываться Ce^{4+} , который переходит в Ce^{3+} . В сурьмяносодержащих стеклах при термообработке облученного стекла реакция (4) не идет. Сурьма способна удерживать электрон вплоть до температуры T_g [3]. При температуре выше T_g сурьма может освобождать электрон, который, в свою очередь, перезахватывается ионом серебра с последующей сегрегацией и образованием коллоидного серебра. Эти реакции могут быть записаны следующим образом



Заключение

Синтезированы фоточувствительные стекла с переменным содержанием сурьмы от 0 до 0,45 мол.%. Показано, что коэффициент поглощения коллоидного серебра симбатно изменяется с содержанием сурьмы в стекле, а при отсутствии сурьмы в стекле коллоидное серебро в процессе термообработки не образуется. Предложен механизм влияния сурьмы на процесс фототермоиндуцированного образования коллоидного серебра.

Литература

1. Tominaga J. The application of silver oxide thin films to plasmon photonic devices// J. Phys.: Condens. Matter. – 15. – 2003. – P.1101–1122.
2. Stookey S. D., Beat G. K., Pierson J. E. Full-color photosensitive glass I J. // Appl. Phys. – 1978. – V. 49. – №10. – P. 5114–5123.
3. Никоноров Н.В., Паньшева Е.И., Туниманова И.В., Чухарев А.В. Влияние состава стекла на изменение показателя преломления при фототермоиндуцированной кристаллизации // Физ. и хим. стекла. – 2001. – Т.27. – № 3. – С. 365–376.
4. Житников П.А., Перегуд Д.Н. ЭПР исследования Ag и Ag₂ центров в фосфатном стекле.// ФТТ. – 1975. – 17. – №6. – С.1655–1660.
5. Паньшева Е.И., Соловьева Н.Д., Туниманова И.В. Взаимодействие УФ и у-излучения с мультихромным стеклом // Физ. и хим. стекла. – 1993. – Т. 19. – № 1. – С. 109–116.
6. Paje S.E., Garcia M.A., Villegas M.A., Plopis J. Optical properties of silver ion-exchangeed antimony doped glass. // J.Non-Cryst. Solid. – 2000. – P.128–136.

РЕАЛИЗАЦИЯ ОДНОКАНАЛЬНОГО СЧЕТА ФОТОНОВ ДЛЯ РАМАНОВСКОГО НАНОСПЕКТРОМЕТРА

П.С. Парфенов

Научный руководитель – д.ф.-м.н., профессор А.В. Баранов

Рассматриваются вопросы реализации методики регистрации излучения на основе одноканального счета фотонов, предназначенной для использования в рамановском наноспектрометре. Исследуются особенности приема и усиления сигнала фотоумножителя в режиме счета фотонов, а также алгоритмы программного управления процессом получения спектра.

Введение

Ближнепольная рамановская спектроскопия является передовым методом контроля с пространственным разрешением нанометрового масштаба таких параметров наноструктур, как размер, форма и взаимное расположение в пространстве, а также химический состав, внутренняя структура и наличие локальных механических напряжений. Наиболее распространенный способ регистрации рамановских спектров, применяемый в спектроскопии – с помощью ПЗС-линейки, что позволяет записывать спектр сразу на определенном интервале длин волн. По такой системе, в частности, построен монохроматор SpectraPro (фирмы Acton Research Corp.), с помощью которого в разработанном нами рамановском наноспектрометре ближнего поля и производится запись спектра.

Однако для ряда задач регистрация с помощью ПЗС-линейки оказывается неподходящей. Во-первых, несмотря на высокую квантовую эффективность и значительное снижение собственных шумов при охлаждении, регистрация сигналов низкого уровня зачастую затруднена из-за природы шумов ПЗС-приемников. Во-вторых, при проведении одноканальной регистрации сигнала из-за малых размеров приемного элемента чувствительность оказывается недостаточной, а принцип функционирования системы считывания затрудняет реализацию на основе ПЗС-линейки синхронного детектирования с модуляцией расстояния «зонд–образец».

Для подобных задач применяются такие приемники, как фотоумножители и лавинные фотодиоды, позволяющие реализовать одноканальное детектирование с высокой чувствительностью и высоким быстродействием, хотя и с меньшей производительностью. Мы разработали для этой цели систему регистрации на основе фотоумножителя, позволяющую при работающем охлаждении регистрировать сигнал, начиная с уровня отдельных фотонов, и проводить исследования на выбранной величине длины волны или величины рамановского сдвига. Для такой системы также разработан алгоритм управления и создано специализированное программное обеспечение для регистрации спектров в одноканальном режиме.

Система приема и усиления оптического сигнала

В ближнепольной рамановской спектроскопии величина полезного сигнала измеряется в отдельных квантах, и в одноканальном режиме для их регистрации применяются фотоэлектронные умножители (ФЭУ) и лавинные фотодиоды (ЛФД). ЛФД ограничены активной зоной 10 мм и усилением от 50 до 200. ФЭУ могут иметь приемную площадку до 50 см, незначительный шум, усиление порядка 10^6 и большой динамический диапазон. Недостатком ФЭУ является меньшая чувствительность – самый чувствительный фотокатод (галлий-арсенид-фосфид) имеет эффективность около 40%, в то время как максимум чувствительности ЛФД и ФД составляет 85%. Также ФЭУ менее чувствительны к длинноволновому излучению. Однако статистическая флуктуация образования и уничтожения электронно-дырочных пар ЛФД создает больший шум, чем диоды ФЭУ [1]. Поэтому в качестве детектора люминесцентного и рамановского из-

лучения нами выбран фотоэлектронный умножитель как прибор, обладающий широким диапазоном и меньшим достижимым уровнем шума.

Помимо темнового сигнала важным параметром ФЭУ является квантовая эффективность счета на данной длине волны. Дело в том, что хорошие с точки зрения темнового сигнала ФЭУ по квантовой эффективности счета могут отличаться друг от друга в 1,5–3 раза. Нами выбран ФЭУ-106 с мультищелочным катодом на основе SbNaKCs и эффективностью счета ~4-9% в диапазоне длин волн люминесценции и рамановского рассеяния 550–650 нм [2]. Этот тип ФЭУ наиболее подходит и по конструктивным соображениям.

Схема испытания и отбраковки ФЭУ

Интенсивность темнового сигнала определяет минимальный регистрируемый полезный сигнал, и приемлемый уровень шума составляет несколько единиц импульсов в секунду. Достичь этого уровня бывает непросто, так как источников шума много – это термоэлектронная эмиссия фотокатода, шум тока утечки между анодом и другими электродами, автоэлектронная эмиссия электродов и пр. [2, 3]. К этому следует добавить эффект старения ФЭУ, в результате которого падает его чувствительность при больших нагрузках и искажаются одноэлектронные характеристики. Отдельные экземпляры ФЭУ одного и того же типа могут значительно отличаться друг от друга по характеристикам и по темновому току даже в пределах одной партии. Поэтому из всех имеющихся в распоряжении следовало выбрать те, которые обладали наименьшими темновыми сигналами.

Для испытания различных экземпляров ФЭУ-106 при работе в одноэлектронном режиме и при охлаждении элементами Пельтье и выбора лучшего из них собран специальный стенд, блок-схема которого показана на рис. 1. В состав стенда входят: светодиод, альтернативно подключаемый к генератору импульсов Г5-15 или к регулируемому источнику питания +3 В; УТФ – устройство для охлаждения фотокатода ФЭУ [4], теплорассеивающий коллектор которого охлаждается проточной водой; высоковольтный блок питания со ступенчатой регулировкой от 0,5 до 2,5 кВ для питания делителя напряжения ФЭУ; эмиттерный повторитель для согласования выходного сигнала ФЭУ с входом предусилителя SR445A, и счетчик фотонов SR400 [5].

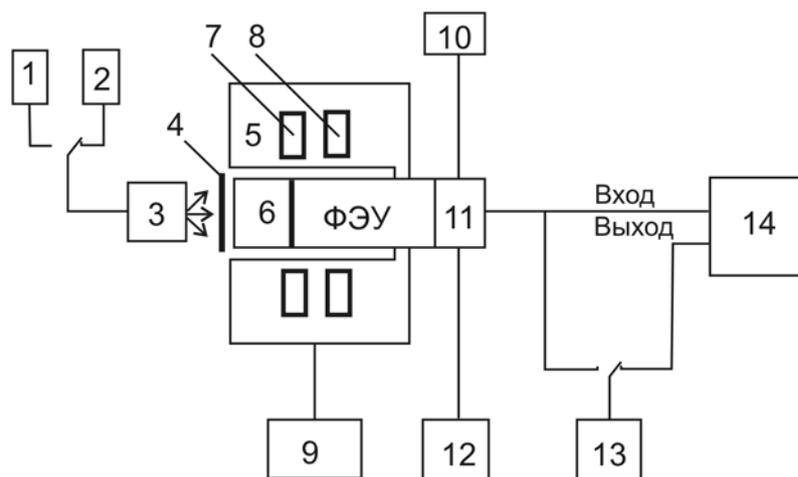


Рис. 1. Блок-схема стенда для испытаний и отбраковки ФЭУ: 1 – генератор Г5-15; 2 – источник питания светодиода +3 В; 3 – светодиод 3ЛЗ41(В); 4 – светоизолирующий затвор; 5 – термоблок для охлаждения фотокатода; 6 – вакуумный объем, обеспечивающий отсутствие запотевания стекла; 7 – теплопоглощающий коллектор; 8 – теплорассеивающий коллектор; 9 – блок питания; 10 – высоковольтный блок питания УБПВ-1; 11 – делитель напряжения, эмиттерный повторитель (ЭП); 12 – блок питания для ЭП; 13 – контрольный осциллограф С1-67; 14 – счетчик фотонов (с предусилителем)

Всего испытано 5 шт. ФЭУ-106. Все блоки и электрические элементы стенда включались заранее за 1–2 часа до начала испытаний данного экземпляра ФЭУ. Сам ФЭУ устанавливался на стенд накануне дня испытаний и надежно изолировался от света с тем, чтобы его фотокатод успел «отдохнуть» от полученных случайных засветок.

Из испытанных экземпляров только один ФЭУ (№ 1666) имеет хорошую счетную характеристику шумовых импульсов при постоянном уровне, соответствующую классическим кривым счетных характеристик ФЭУ [2, 3]. Отчетливо выраженное плато счетной характеристики позволяет выбрать рабочее напряжение для ФЭУ. Оно в данном случае равно 2,0 кВ.

При испытаниях совместно с УТФ (устройство, термоохлаждающее ФЭУ, [4]) отобранный экземпляр ФЭУ также показал наилучшие результаты. Если остальные ФЭУ давали коэффициент уменьшения шума K , т.е. отношение $N_{ш}$ при комнатной температуре к $N_{ш}$ при охлаждении фотокатода, равный 6–10, то экземпляр под №1666 дал $K = 70–100$. Как известно, количество поглощаемой теплоты (или выделяемой на другом коллекторе термобатареи) прямо пропорционально I – силе тока, проходящего через коллектор. Казалось бы, чем больше I , тем больше охлаждение и тем меньше шум ФЭУ. Однако по ряду причин не рекомендуется охлаждать ФЭУ ниже -20°C [2], кроме того, большой ток становится источником лишней теплоты и мешает охлаждению. Поскольку в имеющемся блоке УТФ нет датчиков температур, то ток подобран по максимальному уменьшению шума ($K \cong 100$). Соответствующая оптимальная величина I оказалась равной 2,6 А. Величина K может меняться от измерения к измерению приблизительно на 10% в ту или иную сторону из-за статистической природы шума и внешнего воздействия.

Усиление и регистрация сигнала ФЭУ

Заклученный в охлаждающий корпус ФЭУ мы подключили на выход боковой щели монохроматора Acton SpectraPro-2558, оснащенного голографическим notch-фильтром (режекторный фильтр высокой добротности). Сам монохроматор установлен на оптическом столе, а к его фронтальной выходной щели подключена ПЗС-камера Andor iDus DU401U.

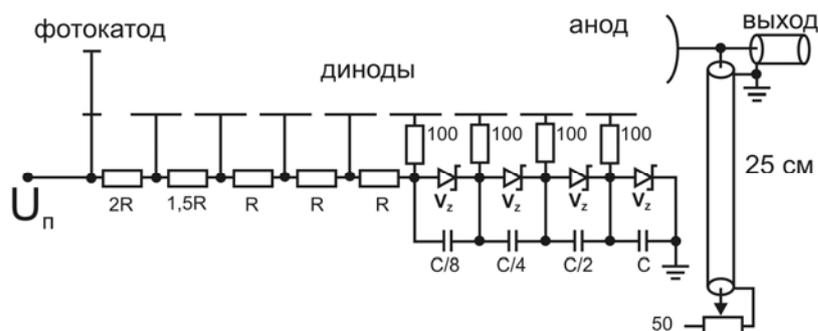


Рис. 2. Схема принципиальная подключения ФЭУ к предусилителю. $R = 180$ кОм, $C = 0,05$ мкФ

Сигнал с выхода эмиттерного повторителя ФЭУ подается на вход предусилителя SR445A, и далее – на счетчик фотонов SR400 [5]. Для счетчика SR400 критична не амплитуда импульса, а время нарастания фронта импульса – для надежного переключения триггера оно должно составлять несколько наносекунд. Для этого ФЭУ подключен по специальной согласующей схеме (рис. 2), рекомендованной изготовителем счетчика [5]. При напряжении питания ФЭУ 1900 В максимальная амплитуда импульсов на выходе повторителя составляет 3 мВ. Счетчик фотонов рассчитан на регистрацию импульсов от 2 до 300 мВ, но рекомендуемая величина – 100 мВ, поэтому для усиления задействованы

два каскада усилителя SR445A (коэффициент усиления каждого каскада $K_{yn}=5$). Анализ амплитудного распределения импульсов показал, что значение уровня дискриминации должно составлять -50 мВ. В этом случае при комнатной температуре регистрируется ≈ 80 имп/с, а при включенной системе охлаждения – единицы импульсов.

Разработанная система приема сигнала позволяет при работающем охлаждении регистрировать сигнал, начиная с уровня отдельных фотонов, и проводить исследования на выбранной величине длины волны или величины рамановского сдвига.

Алгоритм программного управления

Для согласования работы монохроматора и счетчика фотонов мы разработали алгоритм управления и специализированное программное обеспечение (ПО), позволяющие в одноканальном режиме регистрировать спектры рамановского рассеяния или люминесценции. Получение спектров с помощью ФЭУ возможно при реализации пошагового поворота решеток монохроматора ARC SpectraPro-2500i в заданном интервале длин волн и синхронизации работы счетчика фотонов SR400 с перемещением решеток.

Задача разработки ПО возникла вследствие того, что имеющееся штатное ПО монохроматора создано для работы с ПЗС-линейкой и не предусматривает ни сканирующий режим работы, ни синхронизацию работы с другими блоками, а штатное программное обеспечение для счетчика фотонов отсутствует. Однако как для счетчика фотонов, так и для монохроматора предусмотрена возможность управления от компьютера через интерфейс RS-232 и предоставлена документация по наборам используемых команд управления [5, 6]. Кроме того, драйверы монохроматора предоставляют возможность управления через виртуальный последовательный COM-порт, физически реализованный через соединение USB – это обеспечивает большую скорость передачи данных, чем через COM-порт, но оставляет прежним весь набор команд управления монохроматора.

Возможно два режима сканирования спектрального диапазона – пошаговый поворот решетки со счетом сигнала в каждой точке и непрерывный поворот решетки с параллельным счетом сигнала. Преимуществом непрерывного сканирования является меньшее время, требуемое для получения спектра, но мы реализовали первый вариант, так как при пошаговом режиме проще реализовывать синхронизацию.

Для получения спектра в одной точке необходимо выполнить следующие действия (рис. 3, а):

1. последовательно вывести на выходную щель монохроматора излучение на определенных длинах волн в заданном диапазоне;
2. поточно сохранить значение сигнала для этих длин волн;
3. вывести на экран в виде графика и сохранить в виде файла.

Таков алгоритм в первом приближении, рассмотрим его подробнее. При запуске выполняется очистка массива отсчетов, освобождая место для новых данных. Далее осуществляется поворот решетки монохроматора в положение первого отсчета, счетчик переводится в режим единичного счета, приборы получают обновленные параметры работы, вычисляется время цикла, требуемое для получения одного отсчета (складывается из общего времени выборки, «отдыха» счетчика и поворота решетки), запускается счетчик и включается таймер – т.е. запускается цикл опроса.

Циклически, по прерыванию от таймера, считываются данные последнего отсчета (которые заносятся в массив отсчетов), вызывается перерисовка экрана, проверяется достижение верхней границы диапазона (при достижении таймер выключается), решетка монохроматора поворачивается в следующее положение (а значения длины волны и сдвига заносятся в соответствующие массивы), перезапускается счетчик.

Общий алгоритм работы программы (рис. 3, б) включает в себя дополнительные процедуры: создание классов окна и приборов при запуске программы, установки пользователем требуемых параметров, подключения приборов, вывода полученных данных на экран и сохранения в файл, освобождения ресурсов, а также работы с ранее записанными данными без подключения приборов.

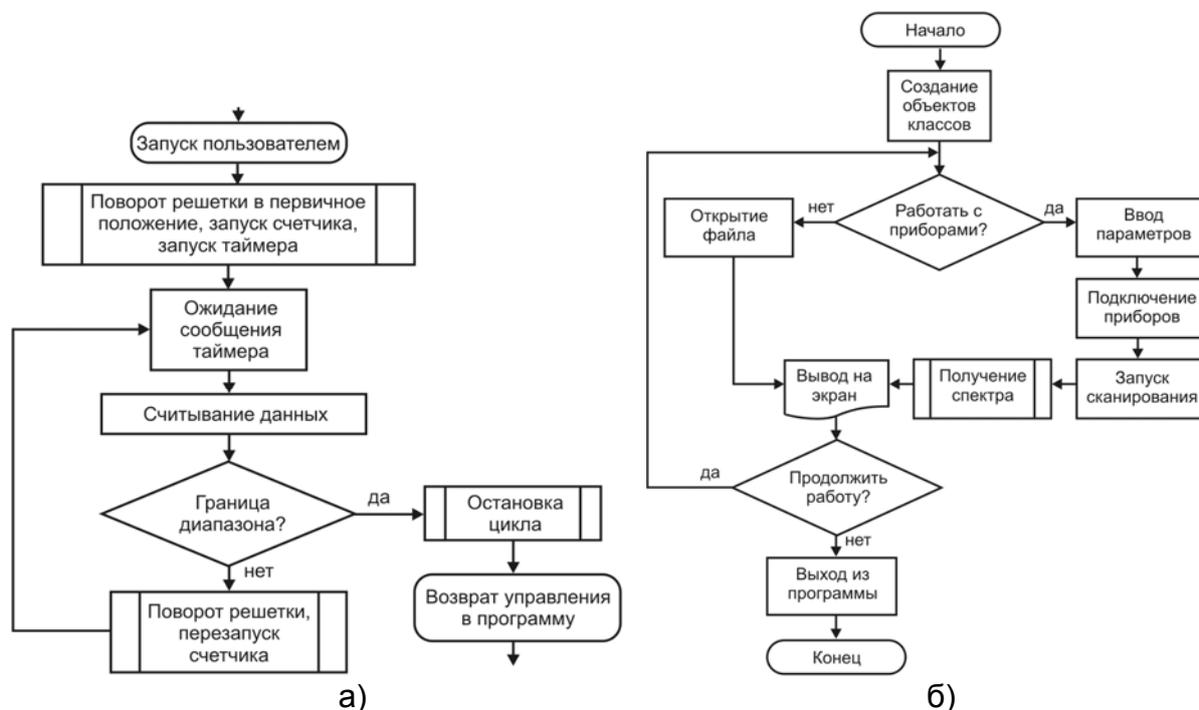


Рис. 3. Функциональная схема работы подпрограммы получения спектра и программы целиком

Для создания управляющей программы «FrameWork» мы выбрали среду разработки Microsoft Visual C++ 6.0 с использованием библиотеки MFC, упрощающей программирование оконных программ под Windows [7]. Программные модули счетчика фотонов SR400 и монохроматора SP2500i реализованы в виде классов SR_COUNTER и ARC_MONOCHROMATOR соответственно (рис. 4). При этом начальная инициализация данных классов производится конструкторами при запуске программы, а непосредственное установление/разрыв связи с приборами производится методами Open/Close. Метод Open обоих классов осуществляет также инициализацию приборов, передавая в них заданные пользователем параметры. Методы этих классов можно условно разделить на две группы – интерфейсные функции, обеспечивающие доступ к данным классов (Get* и Set*– для чтения и записи данных соответственно), и управляющие функции, эквивалентные командам управления приборов.

В уже существующий класс рабочей области окна CChildView помещены не только операции вывода данных на экран в графическом виде (эти операции размещены в функции прорисовки рабочей области OnPaint). Там размещены и функции, отвечающие за работу с приборами (через обращение к методам классов приборов), организацию циклической работы по прерываниям, обработку полученной информации и обработку сообщений пользователя. С одной стороны, из-за этого класс рабочей области окна оказался заполнен функциями, не связанными непосредственно с задачами прорисовки, с другой – разнесение функций по разным классам привело бы к разрастанию количества интерфейсных функций доступа, поскольку работа идет с одним и тем же массивом данных.

Методы работы с файлами, выбора пользователем режима работы приборов (с вызовом методов класса CChildView) и вызова дочерних окон размещены в классе главного окна CMainFrame.

Для хранения данных отсчетов (количества импульсов, длины волны отсчета и величины сдвига) мы использовали три динамических массива класса CArray (на базе шаблонов): m_SRdata, m_ARCwavelength и m_ARCshift размерностью long, double и float соответственно. Функциональная схема взаимодействия объектов программы показана на рис. 4.

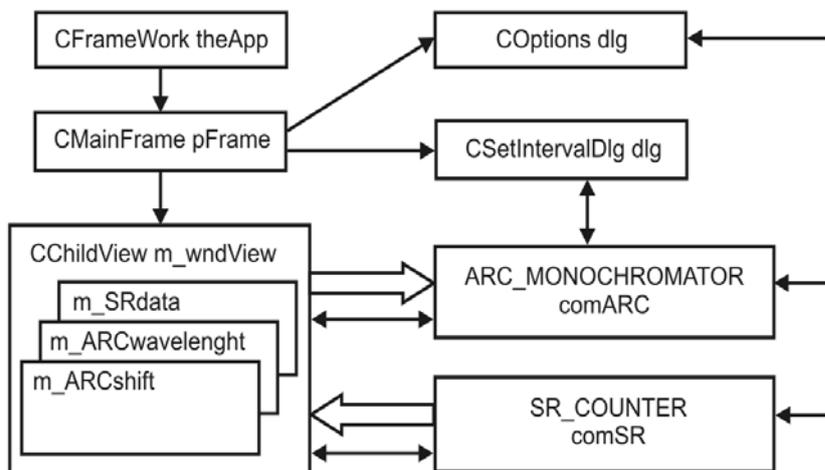


Рис. 4. Функциональная схема взаимодействия классов программы и массивов данных

Для вывода данных на экран мы использовали два основных режима отображения по оси абсцисс – в нанометрах и обратных сантиметрах, и один дополнительный – в порядковых номерах отсчета. Основные режимы предназначены для отображения записываемого спектра, дополнительный режим – для записи сигнала во времени в режиме отладки. Поскольку массиву отсчета соответствуют также и массивы длин волн и сдвига, то при выводе требуется только провести масштабирование графика по размерам экрана.

Для сохранения данных в файл мы выбрали формат хранения отсчетов как отдельных строк в следующем виде (длина волны, величина сдвига и число отсчетов):

nm	cm-1	counts
652.004	467.95	8
652.008	468.044	65
652.012	468.138	83

При чтении файла производится преобразование данных из строкового вида в числовой, с занесением величин в соответствующие массивы.

Для обмена данными с приборами мы задействовали синхронный режим, хотя предпочтительнее выглядит асинхронный режим с использованием прерываний. Оба прибора отвечают на запросы практически мгновенно, и в большинстве случаев задержкой на ответы можно пренебречь. Работа по прерываниям однозначно предпочтительна при ожидании окончания поворота решетки или процесса выборки, но мы ее не реализовали из-за большей сложности. Вместо этого при расчете периода ожидания учитываются как заранее измеренное время, требуемое на поворот решетки, так и заданное пользователем время выборки. Также введена проверка на случай приема ответа «данные не готовы» – в этом случае запрос повторяется.

Счетчик фотонов обеспечивает точность задания времени выборки 0,1 мкс (частота дискретизации 10 МГц) и точность установки уровня дискриминации 0,5 мВ. Ввиду

ненужности такой точности, а также для упрощения расчетов мы установили шаг времени выборки 1 мс и шаг уровня дискриминации 1 мВ.

Внешний вид окна программы показан на рис. 5.

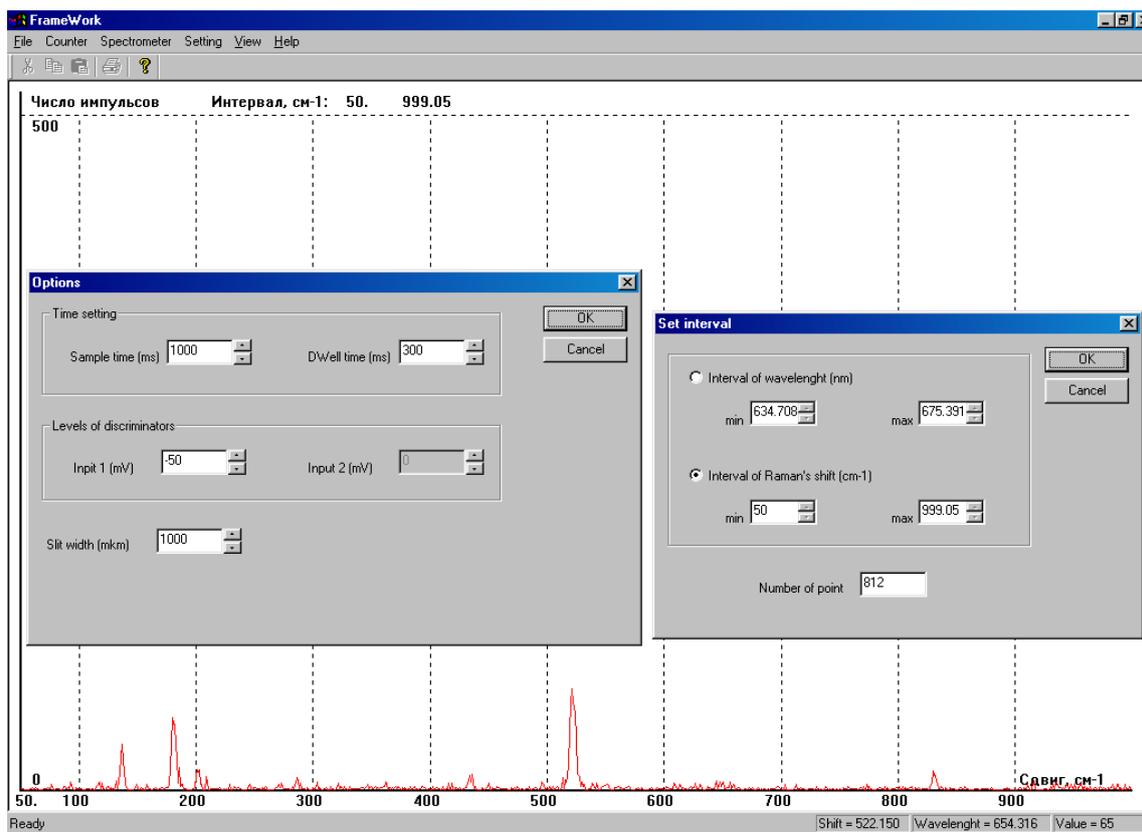


Рис. 5. Окно программы с рамановским спектром кремния и окнами настройки параметров

Разработанный алгоритм управления и созданная на его основе управляющая программа позволяют согласовывать работу счетчика фотонов и монохроматора и получать спектры излучения. Возможно дальнейшее наращивание модулей, например, для реализации функций записи двумерного распределения сигнала по исследуемому участку.

Заключение

Рассмотрена реализация методики регистрации излучения на основе одноканального счета фотонов, предназначенной для использования в рамановском наноспектрометре. На основе проведенного исследования особенностей приема и усиления сигнала ФЭУ в режиме счета фотонов предложен и оптимизирован алгоритм управления записью спектра.

Выражается благодарность С.А. Топоркову за помощь в работе и подготовке этого материала.

Литература

1. Kenneth J. Kaufmann. Light Levels and Noise Guide: Detector Choices // Photonics Spectra. – July 2000. – P. 149–153.

2. Перцев А.Н., Писаревский А.Н. Одноэлектронные характеристики ФЭУ и их применение. – М.: Атомиздат, 1971. – 78 с.
3. Анисимова И.И., Глуховской Б.М. Фотоэлектронные умножители. – М.: Советское радио, 1974. – 64 с.
4. Устройство УТФ. Техническое описание и инструкция по эксплуатации ПУ1.472.030 ТО. 1982. – 19 с.
5. Model SR400. Gated photon counter. Revision 2.5. – Stanford Research System, 2005. – р. 96.
6. Operation instructions Acton Research Corporation SpectraPro-2500i. Rev. 3.08.29. – Acton Research Corporation. – Р. 14.
7. Олафсен Ю., Скрайбнер К., Уайт К. Visual C++ и MFC. Энциклопедия пользователя. – Киев: Диасофт, 2000. – 720 с.
8. Титов О. Работа с коммуникационными портами (COM и LPT) в программах для Win32 [Электронный ресурс]. – Режим доступа: <http://www.realcoding.net/article/view/2416>, свободный.

ТЕХНИКА МИКРО-РАМАНОВСКОГО АНАЛИЗА МАТЕРИАЛОВ

С.А. Черевков

Научный руководитель – д.ф.-м.н., профессор А.В. Баранов

Продемонстрированы возможности оригинальной микроспектральной техники для рамановского и люминесцентного анализа наноструктурированных объектов с дифракционным пространственным разрешением и создания двумерных рамановских изображений объектов.

Введение

Важнейшей проблемой наноиндустрии является развитие неразрушающих методов контроля параметров наноструктур и наноструктурированных материалов, таких как химический состав, внутренняя структура и наличие локальных механических напряжений с достаточным пространственным разрешением. Спектроскопия рамановского или комбинационного рассеяния света является одним из широко используемых методов исследования и характеристики различных материалов. Она позволяет получать информацию о химическом составе, структуре и механических напряжениях без разрушения их структуры.

Химический состав, внутренняя структура и наличие локальных механических напряжений могут быть определены с использованием техники просвечивающей электронной микроскопии и дифракции рентгеновских лучей. Однако эти методы требуют вакуумной техники, долговременной подготовки измерений и часто вызывают разрушение объектов исследования. Аналогичную информацию можно получить при использовании оптической спектроскопии рамановского рассеяния, в частности с помощью микро-рамановской техники, основная деталь которой – микроскоп. Он применяется для визуализации и фокусирования в нужную точку образца, тем самым возбуждая отдельные точки и регистрируя спектр с дифракционным пространственным разрешением, вплоть до 200 нм, за счет хорошей фокусировки и расширения пучка для полного заполнения апертуры микрообъектива. При этом оказывается возможным получать как оптические, так и рамановские изображения исследуемого объекта и на их основе создавать двумерные карты химического состава объекта. Создание таких карт является важной задачей характеристики наноструктур.

В данной работе обсуждается оригинальная микроспектральная техника, которую предполагается использовать для оперативного (несколько десятков минут) получения двумерных рамановских изображений наноструктурированных объектов с дифракционным пространственным разрешением.

Основная часть

Поскольку сигналы рамановского рассеяния малы [1], очень важно для их регистрации использовать светосильный спектральный прибор, не теряя спектрального разрешения, так как полосы рамановского рассеяния значительно уже полос люминесценции. Этого можно добиться с помощью двойного монохроматора с дифракционными решетками с большой плотностью штрихов и большим фокусным расстоянием. В этом случае, однако, возникает проблема со светосилой прибора, определяющей чувствительность измерений. Получение рамановского изображения, для которого необходимо XY сканирование образца с получением рамановского спектра в каждой его точке, требует в данной схеме очень большого времени (до нескольких десятков часов).

Более оптимальным вариантом является применение одиночного монохроматора и использование notch-фильтра для подавления рассеянного лазерного излучения. Для ускорения процедуры регистрации спектров вместо монохроматора с одноканальной регистрацией с помощью ФЭУ [2, 3] современные установки используют спектрографы

с многоканальной системой регистрации, которая позволяет сразу регистрировать достаточно большой участок спектра. В этом случае, однако, получение рамановских изображений, их анализ и построение карт химического состава сталкиваются с трудностями другого рода. Поскольку каждой точке образца здесь сопоставляется спектр, содержащий большое количество полос разной интенсивности, то его анализ требует сложных и долговременных процедур.

Цель работы. Оптическая схема. Основные блоки прибора

Целью работы является оптимизация микро-рамановского спектрометра для получения за разумное время рамановского изображения исследуемого объекта и на его основе – создания карт химического состава с 1:1 соответствием с видимым изображением объекта, получаемым с помощью видеокамеры. Для ускорения процедуры получения рамановских изображений и построения карт химического состава объектов на их основе предложено использовать спектральный прибор, который может работать как спектрограф с многоканальной системой регистрации (ПЗС-камера) и как монохроматор с одноканальной системой регистрации (ФЭУ в режиме счета одиночных фотонов). Использование двух типов детекторов позволяет оптимальным образом организовать получение двумерных рамановских изображений исследуемого объекта. Для этого на первом этапе с помощью многоканального детектора регистрируется спектр рассеяния образца, используемый для идентификации его химического состава и определения наиболее интенсивных характерных полос компонент состава, далее спектрограф переключается в режим монохроматора для одноканальной регистрации наиболее интенсивной линии в спектре с помощью ФЭУ, и производится XY сканирование образца для получения двумерного рамановского изображения. Важно, что в случае одноканальной регистрации появляется возможность использования техники синхронного детектирования, что дополнительно повышает чувствительность измерений.

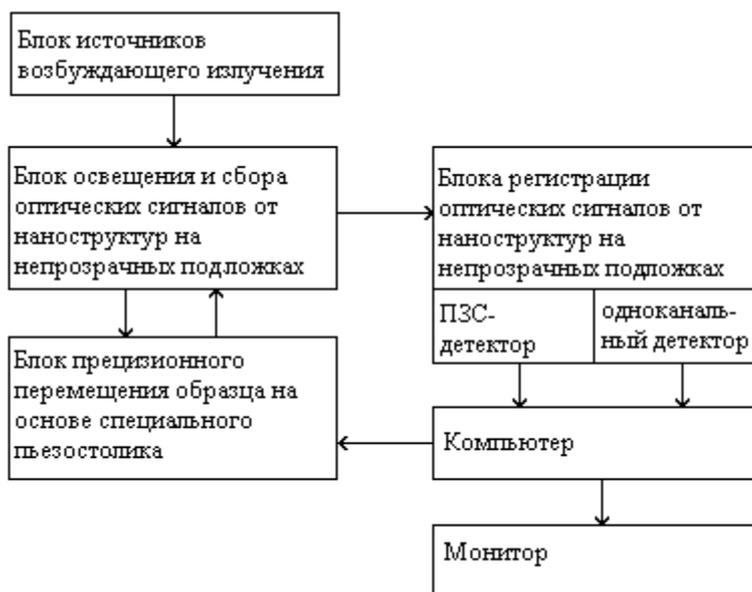


Рис. 1. Блок схема микро-рамановского спектрометра

Блок-схема микро-рамановского спектрометра показана на рис. 1. Он состоит из следующих основных блоков:

- блок источников возбуждающего излучения (YAG:Nd и He-Ne лазеры с длиной волны 532 и 632,8 нм, соответственно) с расширителем пучка;

- блок освещения и сбора оптических сигналов от наноструктур на непрозрачных подложках (на основе люминесцентного микроскопа «Люам И-3, ЛОМО»);
- блок прецизионного перемещения образца на основе специального пьезостолика, контролируемого компьютером, с шагом перемещения 5 нм (3D Flatscanner, Nanonics Imaging Ltd. ISRAEL);
- блок регистрации оптических сигналов от наноструктур на непрозрачных подложках, использующий спектрограф/монохроматор «Acton SP-2558», оборудованный охлаждаемыми многоканальным ПЗС-детектором «Andor iDus DU401A» и одноканальным детектором ФЭУ-79, работающим в режиме счета фотонов

Оптическая схема спектрометра с дифракционным пространственным разрешением и спектральным разрешением 5 см^{-1} представлена на рис. 2.

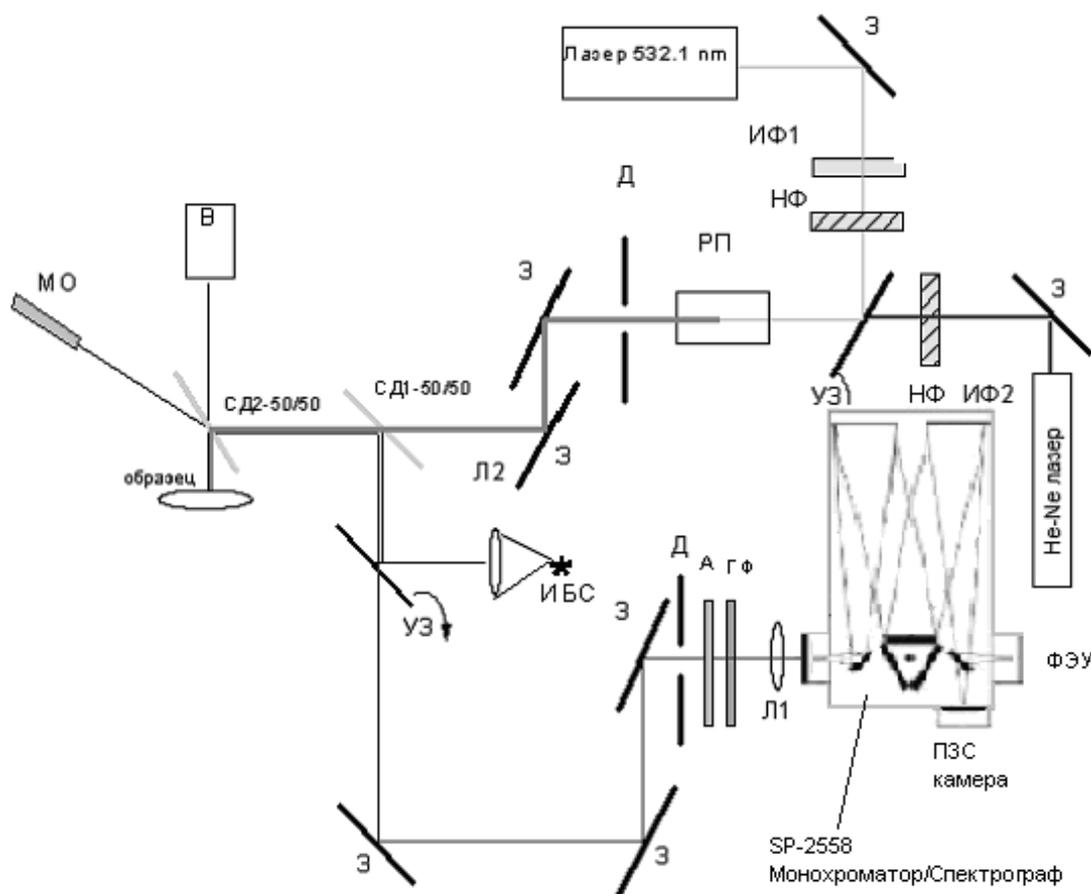


Рис. 2. ИФ1 и ИФ2 – интерференционные фильтры на 532 нм и 638,8 нм, НФ – нейтральный фильтр переменной плотности, СД–О/П – светоделитель с отражением 0% и пропусканием П%, РП – расширитель светового пучка, ИБС – источник белого света, ГФ – узкополосный голографический фильтр (ночь-фильтр), З – зеркало, УЗ – удаляемое зеркало, Д – диафрагма, А – анализатор поляризации, Л – линза, В – видеочкамера, О – объектив, ПЗС – многоканальное регистрирующее устройство, МО – монокуляр для визуального контроля образца, Лазер 532,1 нм и He-Ne лазер – источники возбуждающего излучения

Для оценки уровня рассеянного света и отношения сигнал/шум, а также оптимизации настроек прибора при экспериментальной регистрации спектров микрорамановского рассеяния проведено тестирование прибора, которое заключалось в регистрации спектров рамановского рассеяния монокристаллическими пластинами Si и Ge (MTI Corporation, USA), а также порошка многостенных углеродных нанотрубок (Multiwall

Carbon Nanotube, MWCNT), характерные рамановские спектры которых хорошо известны.

Спектры возбуждались излучением YAG:Nd и He-Ne лазеров с длинами волн 532 и 632,8 нм, соответственно, и мощностью на образце не более 2 мВт. Возбуждающее излучение фокусировалось микрообъективом 50× и численной апертурой 0.65. Оптимизация фокусировки света на образец производится при помощи видеокамеры. Для достижения минимального размера пятна фокусировки лазерный пучок расширялся с помощью телескопа до диаметра, необходимого для полного заполнения апертуры микрообъектива. Большое количество зеркал, используемое в оптической схеме, может вызывать потери сигнала, поэтому в данной установке применялись диэлектрические зеркала с высоким коэффициентом отражения (>97%). Рассеянное излучение собиралось в геометрии обратного рассеяния тем же микрообъективом, коллимировалось, проходило через узкополосный голографический фильтр и фокусировалось на входную щель спектрального прибора. В качестве дисперсионного элемента использовалась дифракционная решетка 1200 штрихов/мм. Спектры регистрировались при спектральном разрешении регистрирующей системы не хуже 5 см^{-1} и времени записи спектра 10 с.

Для снижения уровня рассеянного возбуждающего света – основного источника шумов при регистрации рамановских спектров приборами такого типа – в схеме использованы узкополосные голографические фильтры (ночь-фильтры) фирмы Kaiser, которые эффективно подавляют возбуждающее излучение (532 нм или 632,8 нм) в канале регистрации.

Оптимизация систем возбуждения и сбора излучения для получения максимального пространственного и спектрального разрешений

Расширитель лазерного пучка РП. Для достижения максимального пространственного разрешения необходимо, чтобы возбуждающий лазерный пучок был сфокусирован микрообъективом на образец в пятно с минимальным размером, желательным ограниченным дифракционным пределом. Для достижения минимального размера пятна при фокусировке света конкретным объективом необходимо формирование параллельного лазерного пучка с диаметром, позволяющим максимально заполнить входной зрачок используемого микрообъектива. Для этих целей были использован расширитель пучка (телескоп) лазерного излучения. Размер входного зрачка используемого объектива составляет 7 мм. Характерный диаметр пучков используемых лазеров составляет 0.7 мм для излучения 532 нм (лазер LOS532-200, ООО «Лазеры и оптические системы», Россия) и 1 мм для излучения 632,8 нм (лазер ГН-15-1, «Плазма», Россия). Конструкция телескопа дает возможность прецизионно изменять расстояние между его элементами, позволяя формировать параллельный пучок лазерного излучения из исходных пучков с заметной расходимостью.

Делители световых пучков СД. Оптимизация параметров делителей световых пучков в схеме состоит в нахождении оптимального соотношения коэффициентов отражения O и пропускания Π возбуждающего и регистрируемого излучений для получения максимальной величины анализируемых сигналов (СД1). В то же время должна быть обеспечена возможность надежного визуального наблюдения оптического изображения образца и качества фокусировки возбуждающего излучения (СД2). Оптимизация СД1 достигается при максимальном значении произведения $O \times \Pi$, достигаемом при $O = \Pi = 0,5$. Оптимизация СД2 определяется чувствительностью используемой видеокамеры (SONY-SSC-DC50AP, Япония) и возможностью визуального наблюдения объектов. Сопоставление визуального и телевизионного изображений показало, что для СД2 соотношение $O/\Pi = 0.5/0.5$ является оптимальным.

Микрообъектив. Для достижения максимального пространственного разрешения необходимо, чтобы качество сборки микрообъектива (центровка линз объектива) по-

зволяло достигнуть дифракционного предела фокусировки. Качество сборки может быть проверено непосредственно с использованием оптической схемы прибора. При лазерном освещении микрообъектива, несмотря на просветляющие покрытия линз, имеет место слабое отражение излучения от поверхностей линз, формирующее интерференционную картину, наблюдающуюся в виде концентрических колец визуально или с помощью видеокамеры. В случае плохой центровки линз наблюдается наличие нескольких систем концентрических окружностей с несовпадающими центрами.

Ряд элементов наноспектрометра используется одновременно в системах возбуждения и сбора излучения, например, микрообъектив и часть светоделителей и зеркал (рис. 2), поэтому процедуры выбора и оптимизации параметров этих элементов, описанные в предыдущем разделе, также оптимизируют сбор вторичного излучения образца. Процедура оптимизации положения (нахождение рабочего угла) узкополосного голографического фильтра необходима для подавления упруго рассеянного света и увеличения отношения сигнал/шум и, следовательно, для повышения точности пространственных и спектральных измерений. Остановимся здесь на оптимизации по светосиле и разрешающей способности ввода вторичного излучения в спектрограф.

Оптимизация ввода вторичного излучения в спектральный прибор. Для достижения оптимальной светосилы используемого спектрального прибора необходимо, чтобы телесный угол сбора излучения на щель прибора был равен входному углу спектрометра [7]. Это условие также соответствует достижению максимальной разрешающей способности спектрального прибора, поскольку при этом происходит полное заполнение светом площади дифракционной решетки. В нашем случае сбора света из практически точечного источника можно ограничиться условием равенства плоских углов. Спектрограф ACTON SpectraPro-2558, используемый в схеме наноспектрометра, имеет фокусное расстояние 500 мм и размер дифракционных решеток 68×68 мм, поэтому его входной угол равен $\sim 68/500 = 0.136$. В используемой нами геометрии сбора света вторичное излучение, выходящее из микрообъектива, представляет собой параллельный пучок с диаметром 7 мм. Поэтому оптимальная светосила и разрешение будут реализовываться при фокусировке излучения на щель линзой с фокусным расстоянием $F = 7/0.136 = 51.5$ мм. При использовании такой линзы в области входной щели прибора формируется изображение освещенного микрообъективом участка образца, увеличенное в ~ 10 раз (фокусное расстояние микрообъектива равно 5 мм). При размере пятна < 2 мкм его размер будет составлять < 20 мкм, что меньше ширины входной щели прибора, соответствующей заданному в ТЗ минимальному значению спектрального разрешения 5 см^{-1} при обоих типах дифракционных решеток спектрографа (1200 шт/мм и 2400 шт/мм). Это обстоятельство позволяет нам достигнуть оптимального спектрального разрешения регистрируемых сигналов без дополнительных потерь в их интенсивности.

Регистрация спектров рамановского рассеяния образцов с известными характеристиками

Для определения спектральных параметров прибора и его чувствительности была проведена регистрация спектров рамановского рассеяния образцов с известными характеристиками (пленки Si, Ge, и углеродные нанотрубки). Измерены интенсивности рамановских сигналов и уровни фоновых шумов в спектрах при спектральном разрешении регистрации менее 5 см^{-1} и временем записи 10 с. Продемонстрировано, что для всех тестируемых образцов отношение сигнал/шум не меньше 30.

Спектральное разрешение и чувствительность прибора проиллюстрированы типичными спектрами, приведенными на рис. 3 и 4. На рис. 3 показаны спектры микро-рамановского рассеяния образца монокристаллического кремния, возбуждаемые излучением 532 нм (А, В и Д) и 632,8 нм (Б, Г и Е). Спектры в каждой из серий показывают: А, Б – общий вид спектра, содержащего характерную линию Si 520 см^{-1} ; В, Г – деталь-

ный спектр этой линии, демонстрирующий, что спектральное разрешение прибора лучше, чем 5 см^{-1} ; Д,Е – участок фонового спектра, показывающий уровень шумов.

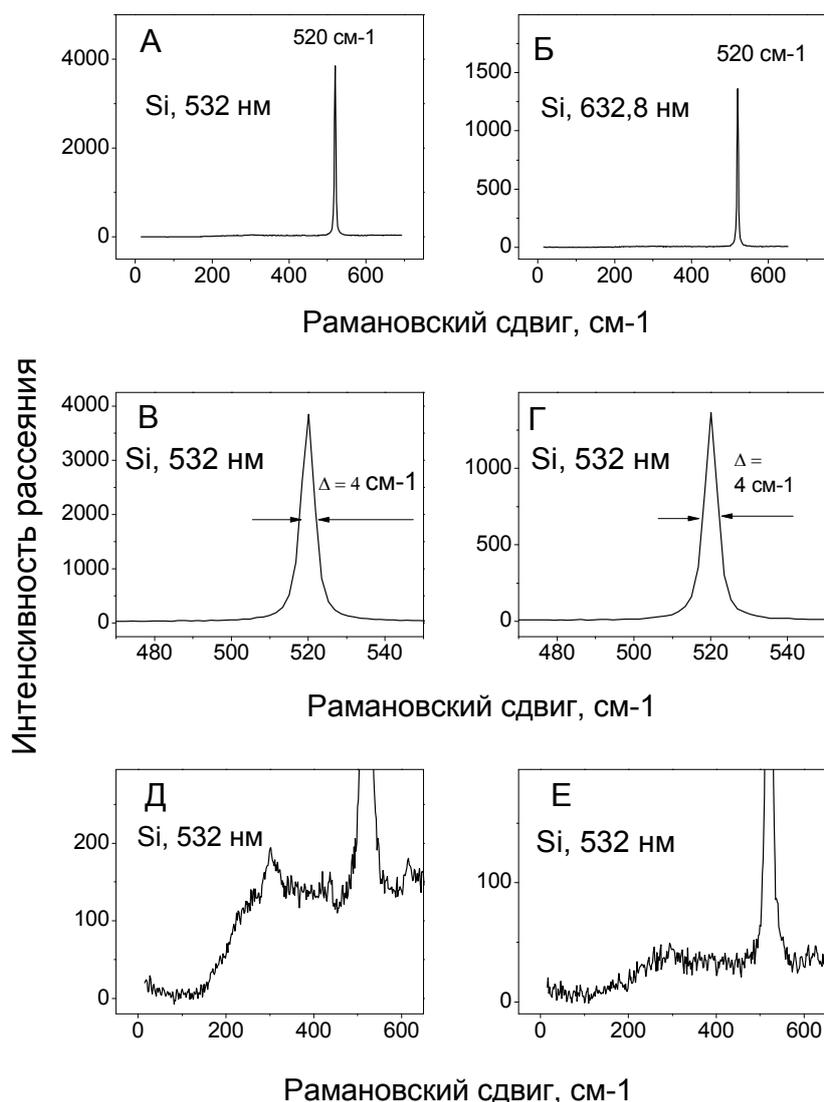


Рис. 3. Спектры микро-рамановского рассеяния образца монокристаллического кремния, возбуждаемые излучением 532 нм (А, В и Д) и 632,8 нм (Б,Г и Е)

На рис. 4 приведены спектры микро-рамановского рассеяния образца монокристаллического германия и порошка многостенных углеродных нанотрубок, возбуждаемые излучением 532 нм (А и В) и 632,8 нм (Б и Е), демонстрирующие общий вид характерных рамановских линий Ge и углерода (305 см^{-1} и 1581 см^{-1} , соответственно). Детальный анализ, аналогичный проведенному для спектров Si, показывает практически такие же значения интенсивности шума.

Результаты тестовых измерений показали, что параметры прибора оптимизированы для регистрации спектров микрорамановского рассеяния макрообразцов с известными характеристиками (Si, Ge, C) при спектральном разрешении не менее 5 см^{-1} , временем записи не более 10 с и отношением сигнал/шум не менее 10.

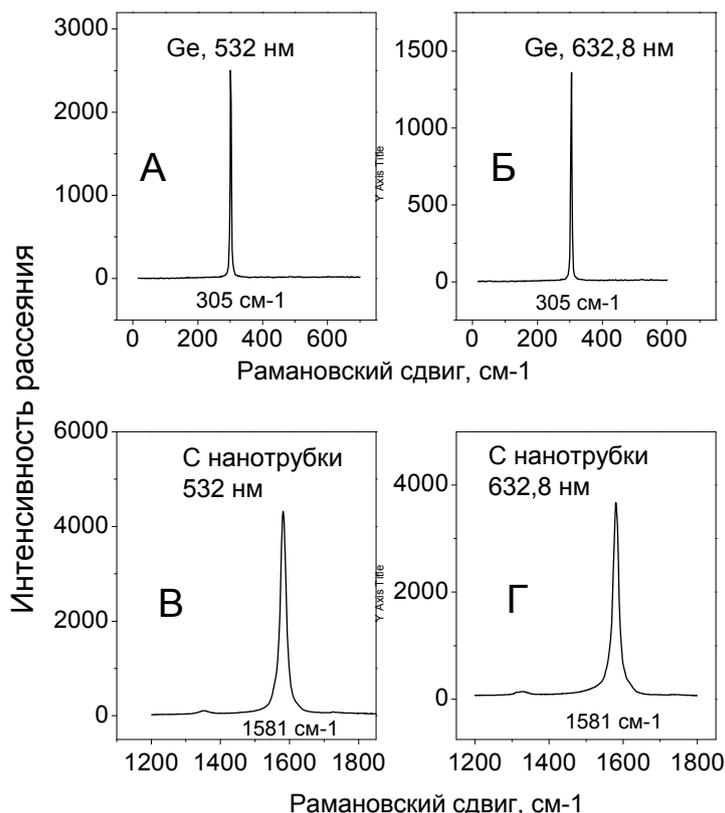


Рис. 4. Спектры микро-рамановского рассеяния образца монокристаллического германия и порошка многостенных углеродных нанотрубок, возбуждаемые излучением 532 нм (А и В) и 632,8 нм (Б и Е)

Проведенные измерения рамановских спектров тех же образцов с использованием одноканальной регистрации с помощью ФЭУ показали, что чувствительность измерений сигналов остается примерно на том же уровне, что и для многоканальной регистрации [8]. Как и ожидалось, время записи того же участка спектра при одноканальной регистрации увеличилось примерно в 200 раз. Однако этот параметр несущественен для предлагаемой методики регистрации рамановских изображений, в которой происходит регистрация интенсивности сигнала с выбранным рамановским сдвигом без сканирования монохроматора.

Проведенные оценки показали, что при достигнутой чувствительности определений для уверенной регистрации рамановского изображения участка объекта состоящего из 512 пространственных точек ($500 \times 500 \text{ мкм}^2$, при размере пятна фокусировки в 1 мкм), необходимо 2100 сек или 0,57 часа. В расчете предполагалось, что объект представляет собой Si/Ge сплав.

Заключение

Проведенное тестирование разрабатываемой микро-спектральной техники показало, что она может быть использована для рамановской спектроскопии и получения рамановских изображений объектов микронного размера за приемлемое время.

Литература

1. Long D.A. The Raman effect. – Wiley, 2002. – 610 p.

2. Перцев А.Н., Писаревский А.Н. Одноэлектронные характеристики ФЭУ и их применение. – М.: Наука, 1971.
3. Анисимова И.И., Глуховской Б.М. Фотоэлектронные умножители. – М. Наука, 1974.
4. Raman Microscopy. Edd by G. Turrell and J. Corset. –Elsewier, 1996. – 453 p.
5. Ландсберг Г.С. Оптика. – М.: Наука, 1976.
6. S. Nie, S.R. Emory. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering // Science. – 1997. – Vol. 275. – P. 1102–1106.
7. N. Hayazawa, Y. Inoue, Z. Sekkat, S. Kawata. Near-field Raman imaging of organic molecules by an apertureless metallic probe scanning optical microscope // J. Chem. Phys. – 2002. – Vol. 117. – №. 3. – P. 1296–1301.
8. L.T. Nieman, G.M. Krampert, and R.E. Martinez. An apertureless near-field scanning optical microscope and its application to surface-enhanced Raman spectroscopy and multiphoton fluorescence imaging // Review of Scientific Instruments. – 2001. – Vol. 72. – №. 3. – P. 1691–1699.

ПЕРЕНОС ЭНЕРГИИ ФОТОВОЗБУЖДЕНИЯ В СИСТЕМАХ КВАНТОВЫХ ТОЧЕК

Е.В. Ушакова

Научный руководитель – д.ф.-м.н., профессор А.В. Баранов

Рассмотрены особенности проявления переноса энергии в системах квантовых точек и, в частности, безызлучательный резонансный перенос энергии (FRET). Обсуждаются методы исследования переноса энергии с использованием временных характеристик затухания люминесценции. Продемонстрированы возможности микролюминесцентной техники на основе прибора MicroTime 100, в том числе и возможность создания 2D изображения эффективности процесса переноса энергии.

Введение

В последнее время большое внимание уделяется изучению наноструктур. Люминесцирующие полупроводниковые нанокристаллы размером в несколько нанометров (квантовые точки, КТ) считаются перспективными для создания нового типа люминесцентных сенсоров или материалов с переносом энергии для систем сбора света.

Развитие микролюминесцентной техники и создание лазеров с пикосекундными импульсами излучения привело к значительному увеличению чувствительности определений, а также возможности измерения не только спектральных, но и временных параметров оптических откликов исследуемых объектов, в частности, к возможности измерения времен затухания люминесценции структур на основе квантовых точек [1], по результатам которых можно судить о релаксации фотовозбужденных носителей в КТ и о процессах передачи энергии в системах с близко расположенными КТ. Одним из наиболее эффективных механизмов передачи энергии в таких системах является безызлучательный перенос энергии – Fluorescence Resonance Energy Transfer (FRET). О наличии такого процесса передачи энергии и о его эффективности можно судить не только по тушению люминесценции КТ-донора и увеличению интенсивности люминесценции КТ-акцептора, но и по изменению их времен затухания люминесценции. В случае FRET характерные времена затухания люминесценции донора уменьшаются, а акцептора увеличиваются. Во многих случаях анализ времен затухания позволяют получить более детальную информацию о взаимодействии КТ и о процессах переноса энергии. Особенно это важно в случае исследования относительно протяженных двумерных структур, где параметры FRET могут меняться от точки к точке, т.е. где имеется необходимость создания 2D изображения эффективности процесса переноса энергии.

В работе подробно анализируются возможности современной микролюминесцентной техники на основе прибора MicroTime 100 при измерении времен жизни люминесценции квантовых точек (0,1–1000 нс) и создания двумерных пространственных карт распределения времен жизни с дифракционным пространственным разрешением. Временные измерения проводятся с использованием техники коррелированного счета одиночных фотонов.

Перенос энергии в системах квантовых точек

Квантовые точки являются одним наиболее исследуемых объектов физики низкоразмерных полупроводниковых гетероструктур. Квантовая точка – это кристалл полупроводника с характерным размером меньшим или порядка боровского радиуса экситона, R_{ex} . Из-за малого размера подобные кристаллы обладают свойствами, отличными от «больших» полупроводниковых кристаллов. Пространственное ограничение движения носителей заряда приводит к квантово-размерному эффекту, выражающемуся в дискретной структуре электронных уровней, из-за чего КТ иногда называют «искусственными атомами».

Особый интерес представляют люминесцирующие квантовые точки, в которых поглощение фотона рождает электрон-дырочные пары, а рекомбинация электронов и дырок вызывает люминесценцию. Одним из уникальных свойств квантовых точек является непрерывное поглощение в широком диапазоне длин волн (рис. 1), что позволяет возбуждать разные квантовые точки светом с одной длиной волны. При этом положение экситонного пика поглощения (крайнего справа на рис. 1) определяется размером нанокристалла. Такие квантовые точки обладают узким и симметричным пиком люминесценции, положение которых также определяется размером КТ (рис. 1).

Кроме того, КТ обладают более высокой яркостью и лучшей фотостабильностью по сравнению с традиционными люминофорами.

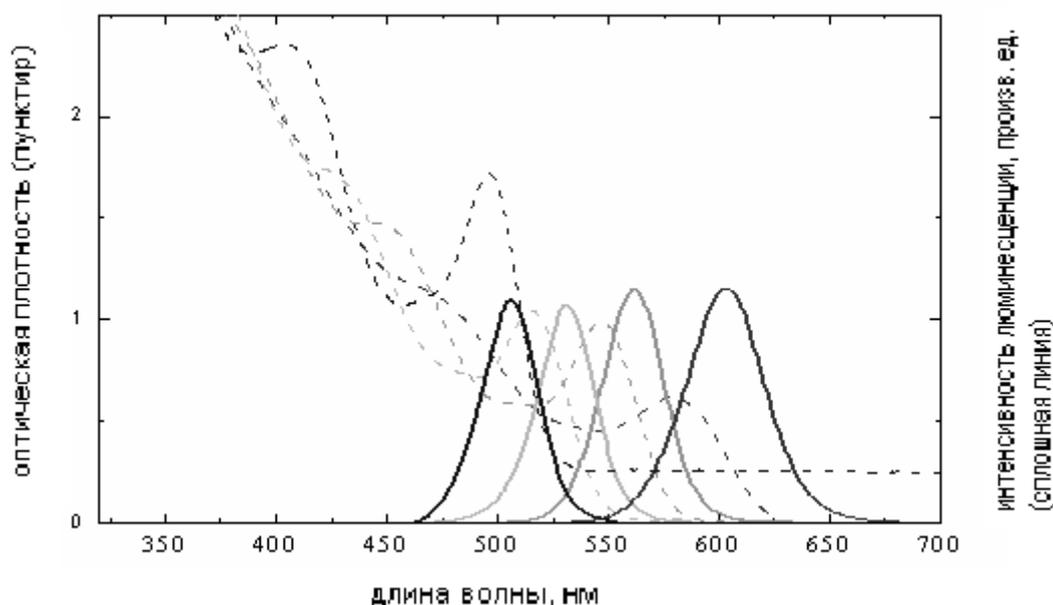


Рис. 1. Спектры поглощения и люминесценции квантовых точек CdSe (разных по размеру)

Перенос энергии – это способность одной компоненты системы (донора) передать энергию возбуждения другой компоненте (акцептору). Одним из механизмов переноса энергии является FRET. Для возможности реализации переноса энергии должны выполняться определенные условия. Одним из основных является условие резонанса между донором и акцептором – спектральное перекрытие полосы люминесценции донора с полосой поглощения акцептора. Процесс передачи энергии должен происходить быстрее, чем процесс излучательного перехода донора из возбужденного состояния в основное. Вторым необходимым условием является близость донора и акцептора. Из формулы (1) видно, что эффективность FRET сильно зависит от расстояния (R) между донором и акцептором.

$$Q_{FRET} = \frac{R_0^6}{R_0^6 + R^6}, \quad (1)$$

где R_0 – ферстеровский радиус. Поэтому передача энергии от донора к акцептору происходит достаточно эффективно при расстояниях (R) порядка 10–100 Å.

Таким образом, донор возвращается в основное состояние без излучения поглощенной энергии, которая, будучи переданной акцептору, высвобождается либо в виде излучения, либо другими путями диссипации с возбужденного уровня акцептора.

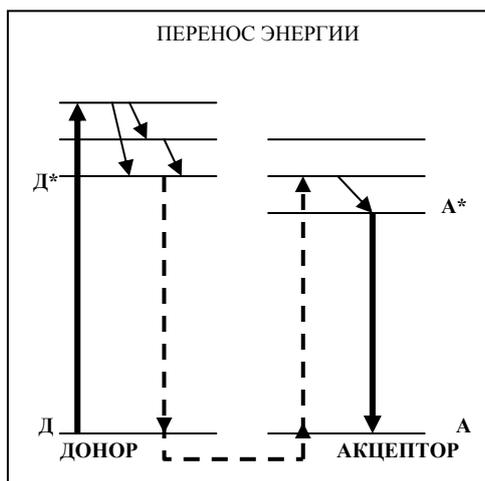


Рис. 2. Схема передачи энергии от донора к акцептору

Этот процесс приводит к тушению люминесценции донора и увеличению интенсивности люминесценции акцептора. Перенос энергии не сопровождается значительными изменениями в спектрах поглощения и излучения донора и акцептора. Схематично процесс FRET иллюстрируется на рис. 2:

- 1) поглощение донором фотона с переходом в возбужденное электронное состояние (Д → Д*);
- 2) колебательная релаксация в нижнее возбужденное электронное состояние донора;
- 3) передача возбуждения от донора к акцептору (А*);
- 4) колебательная релаксация в молекуле акцептора;
- 5) рекомбинация электрон-дырочной пары в акцепторе с испусканием света.

Скорости процессов «2» и «4» обычно велики и равны 10^{11} – 10^{12} с⁻¹. Процесс «4» очень важен в акте безызлучательного переноса энергии, так как при этом система выходит из резонанса и обратный перенос энергии становится невозможным.

В эксперименте обычно измеряются следующие характеристики процесса переноса энергии:

- 1) уменьшение интенсивности люминесценции донора в присутствии акцептора;
- 2) увеличение интенсивности люминесценции акцептора в присутствии донора;
- 3) уменьшение (увеличение) времени затухания люминесценции донора (акцептора) в присутствии акцептора (донора).

Эффективность FRET может быть определена по изменению как интенсивности, так и времени затухания люминесценции донора в присутствии акцептора:

$$Q_{FRET} = 1 - \frac{I_{DA}}{I_D}, \quad Q_{FRET} = 1 - \frac{\tau_{DA}}{\tau_D},$$

где I_{DA} , I_D , τ_{DA} , τ_D – интенсивность и времена затухания люминесценции донора в присутствии и отсутствии акцептора, соответственно.

Полупроводниковые квантовые точки предоставляют уникальные возможности для изучения эффектов передачи энергии фотовозбуждения и создания на их основе люминесцентных сенсоров и новых материалов для сбора и концентрации энергии. Действительно, они эффективно поглощают свет в широком диапазоне длин волн, обладают большим (до 80–85%) квантовым выходом люминесценции и, что очень важно, возможностью оптимизации резонансных условий переноса энергии путем подбора оптимального размера КТ-донора и /или КТ-акцептора.

Во многих случаях, особенно в структурах с квантовыми точками, использование методов измерения люминесценции затруднено, и информация о процессах переноса

энергии можно более точно получить при помощи временного анализа затухания люминесценции.

Измерение и картирование времен жизни люминесценции

Последние достижения в генерации коротких (пикосекундных и субпикосекундных) световых импульсов делают возможным исследование быстрых процессов, происходящих при возбуждении и деактивации электронных возбуждений в молекулах и нанокристаллах. В настоящее время одним из эффективных методов изучения наноструктур при низкой интенсивности люминесценции и временах затухания люминесценции выше единиц пикосекунд является техника коррелированного счета одиночных фотонов (Time-Correlated Single Photon Counting, TCSPC), которая обладает высокой чувствительностью и хорошим временным разрешением люминесцентных измерений. Такую технику измерения еще называют техникой задержанных совпадений, она позволяет измерять времена затухания люминесценции вплоть до нескольких пикосекунд.

Рассмотрим более подробно процедуру измерения временных характеристик люминесценции с помощью данной техники. Для возбуждения люминесценции в этом методе используется последовательность коротких возбуждающих импульсов излучения. Вероятность детектирования люминесценции поддерживается на уровне ниже одного фотона при возбуждении объекта одним импульсом, а частота повторения возбуждающих импульсов делается как можно большей. Используется методика счета отдельных фотонов. При этом измеряется временное распределение вероятности $P(t) dt$ того, что после импульса возбуждения в момент $t = 0$ фотон люминесценции испущен в интервале времени от t до $t+dt$.

Такой метод измерения времен релаксации люминесценции лежит в основе работы прибора MicroTime 100 [6], блок-схема которого приведена на рис. 3.

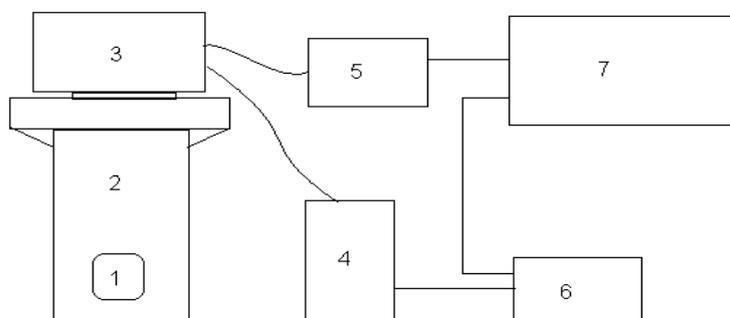


Рис. 3. Схема установки MicroTime 100 (PicoQuant, Германия): 1 – образец, 2 – люминесцентный микроскоп (BX-51), 3 – устройство, управляющее лазерным излучением для сканирования объекта, 4 – диодный пикосекундный импульсный лазер (LDH-P-C-405-B), 5 – приемник излучения (PMA 185-P-M), 6 – Блок питания и контроля лазера (PDL 800-B), 7 – персональный компьютер со встроенной платой для проведения временных измерений TimeHarp 200 и программным обеспечением для обработки данных SPT-2 «SymPhoTime»

Рассмотрим более подробно принцип работы и технические характеристики данного устройства. Установка MicroTime100 основана на базе «прямого» люминесцентного микроскопа. Излучение от диодного лазера (4) с длительностью импульсов менее 70 пс и частотой повторения 40 МГц заводится с помощью оптоволоконного кабеля в микроскоп через устройство сопряжения (2). С помощью этого устройства лазерный луч осуществляет XY-сканирование с заданным временем накопления t , позволяя фиксировать кинетику люминесценции в каждой точке объекта с пространственным раз-

решением 2 мкм. Излучение от образца проходит через дихроичный фильтр и с помощью оптоволокну заводится в приемник излучения (5), который состоит из предусилителя, обтюратора, блока питания и ФЭУ с областью спектральной чувствительности в 300-820 нм. Далее сигналы с устройств (5) и (6) подаются в компьютер, где с помощью встроенной платы TimeHarp 200 производится обработка данных с временным разрешением менее 100 пс. Анализ данных осуществляется с помощью программы SymPhoTime, в которой можно строить двумерные изображения времен затухания люминесценции. В данной установке могут быть использованы импульсные пикосекундные диодные лазеры, излучающие в диапазоне длин волн 375–800 нм.

Возможности данного прибора:

- составление двумерных карт распределения времен жизни люминесценции различных веществ (FLIM – fluorescence Lifetime Imaging);
- построение гистограмм количества фотонов с разными временами затухания;
- дифференцирование различных люминесцентных меток с одинаковыми спектральными свойствами (спектрами поглощения/излучения);
- измерение размеров люминесцирующих областей с точностью до нескольких нанометров;
- количественные исследования различных сред, в которых наблюдается безызлучательный перенос энергии.

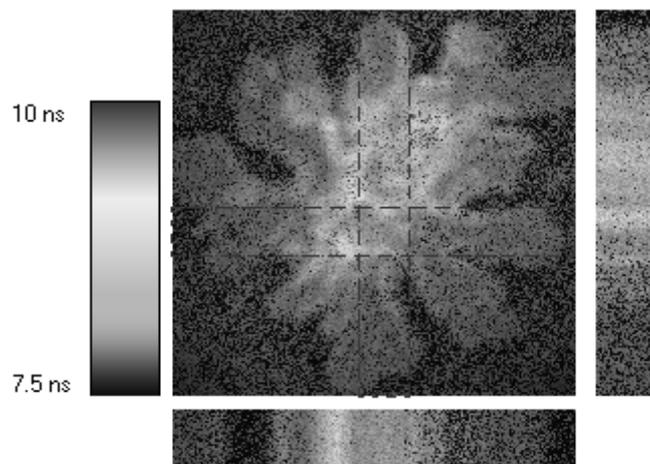


Рис. 4. Двумерное изображение времен затухания люминесценции КТ в дендритоподобной структуре. Возрастание времен от периферии структуры к ее центральным областям связано с переносом энергии

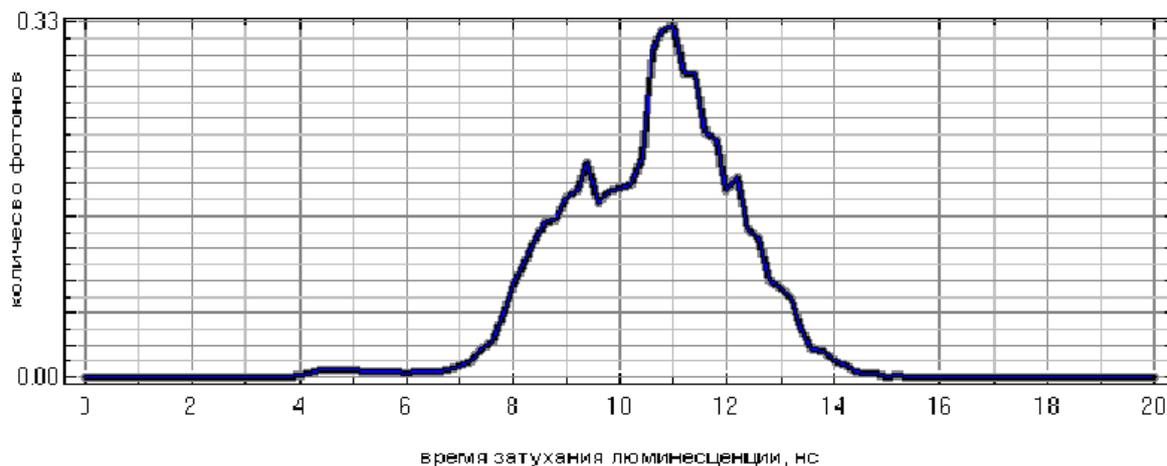


Рис. 5. Гистограмма распределения количества фотонов по временам затухания люминесценции КТ

В качестве иллюстрации возможностей FLIM на рис. 4 показана двумерная карта распределения времен жизни люминесценции квантовых точек CdSe/ZnS, образовавшихся в результате процесса самоорганизации дендритоподобную структуру микронных размеров, в которой имеет место перенос энергии фотовозбуждения от периферийных областей к центру.

На рис. 5 приведен пример гистограммы распределения количества фотонов по временам затухания люминесценции квантовых точек, из анализа которой можно сделать выводы о наличии разной эффективности взаимодействия КТ в разных областях наноструктуры.

Заключение

Применение микролюминесцентных приборов, в основе работы которых лежит техника коррелированного счета фотонов (TCSPC), и применение пикосекундных импульсных лазеров дает возможность с большой точностью изучать временные характеристики безызлучательного переноса энергии фотовозбуждения (FRET). Применение особых устройств сканирования образца или лазерного луча по образцу дает возможность строить двумерные карты распределения времен жизни люминесценции, по которым можно судить об эффективности взаимодействия КТ, процессах релаксации и передачи энергии как во всей наноструктуре, так и в ее локальных областях. Важно, что, используя распределение времен жизни люминесценции в зависимости от координат, можно определить химический состав в каждой точке образца, и на основе этих данных строить двумерные карты химического состава исследуемого объекта.

Такие исследования временных параметров переноса энергии в системах КТ позволяют более детально изучить их уникальные оптические свойства, на основе которых сейчас создаются новые материалы для сбора и передачи энергии, например, солнечные батареи высокой эффективности, и различных видов сенсоров таких как, флуоресцирующие маркеры, фотосенсибилизаторы в медицине, фотодетекторы в ИК-области и др.

Литература

1. Дермтредер В. Лазерная спектроскопия: Основные принципы и техника эксперимента: Пер. с англ./Под ред. И.И. Собельмана. – М.: Наука. Главная редакция физико-математической литературы, 1985. – 608 с, ил.
2. Оптика наноструктур. Под редакцией А.В. Федорова: СПб: Недра, 2005. – 326 с.
3. W.E. Moerner and David P. Fromm. Methods of single-molecule fluorescence spectroscopy and microscopy. // Review of scientific instruments. – 2003. –V. 74. – № 8. – P. 3597–3619.
4. A. Sukhanova, A.V. Baranov, T.S. Perova, J.H.M. Cohen, and I. Nabiev. Controlled Self-Assembly of Nanocrystals into Polycrystalline Fluorescent Dendrites with Energy-Transfer Properties. // Angew. Chemie Int. Ed. – 2006. – 45(13). – 2048–2052.
5. Ермолаев В.Л., Бодунов Е.Н., Свешникова Е.Б., Шахвердов Т.А. Безызлучательный перенос энергии: – Л.: Наука, 1977. – 331 с.
6. PicoQuant GmbH [Электронный ресурс]. – Режим доступа: <http://www.picoquant.com>

САМООРГАНИЗАЦИЯ КВАНТОВЫХ НАНОКРИСТАЛЛОВ CdTe НА ГИДРОФОБНОЙ ПОВЕРХНОСТИ

В.Е. Адрианов, Г.Н. Виноградова

Научный руководитель – д.ф.-м.н., профессор А.В. Баранов

Сообщается о результатах исследования процесса самоорганизации квантовых нанокристаллов CdTe, путем испарения водного коллоидного раствора на гидрофобной и гидрофильной поверхностях.

Введение

Одним из перспективных направлений нанотехнологических исследований является изучение процессов самоорганизации нанокристаллов или индуцированного упорядочения наночастиц на различных твердотельных подложках. Ансамбли наночастиц с высокой степенью упорядочения часто обладают новыми, неожиданными свойствами и являются потенциальными рабочими элементами различных устройств. В этом направлении открываются возможности создания материалов для нового поколения устройств, архитектура которых основывается на способности нанокристаллов образовывать упорядоченные ансамбли в результате термодинамически индуцированных процессов самоорганизации или управляемой внешними воздействиями упорядоченности. Меняя размер, форму, тип поверхностных покрытий и заряд на поверхности нанокристаллов, можно контролировать процесс самоорганизации и фабриковать упорядоченные структуры различного типа. И, что важно, размерная зависимость физических свойств нанокристаллов (НК) позволяет целенаправленно управлять ими, открывая возможности манипуляции параметрами создаваемых ансамблей с улучшенными оптическими свойствами. При резком расширении областей применения нанокристаллов, включающих электронику, оптоэлектронику, оптику, энергетику и другие промышленные области, данные исследования являются исключительно полезными для конструирования элементов устройств, применяемых в этих отраслях.

Ранее самоорганизация нанокристаллов изучалась на примере нанокристаллов CdSe/ZnS на гидрофобных и гидрофильных поверхностях [1, 2]. Были получены такие структуры, как дендриты и упорядоченные гексагональные решетки. Было обнаружено, что дендритные структуры обладают свойством переноса энергии от периферийных областей к центру дендрита. Были определены основные особенности процесса термодинамически индуцированной самоорганизации этих нанокристаллов. Осталось выяснить, является ли обнаруженные особенности общими для нанокристаллов разных типов, или же это специфика нанокристаллов CdSe/ZnS. В данной работе исследовались нанокристаллы CdTe, которые имеют отличающиеся оптические свойства и другой тип молекул на поверхности нанокристаллов. В частности, спектр люминесценции CdTe сдвинут в красную область по сравнению со спектром люминесценции нанокристаллов CdSe/ZnS, что может быть выгодно использовано во многих оптоэлектронных приборах и устройствах.

В настоящей статье сообщается о получении самоорганизованных структур при испарении водного коллоидного раствора полупроводниковых нанокристаллов (квантовых точек, КТ) CdTe на гидрофобных и гидрофильных поверхностях твердого тела. Также описываются исследования закономерностей и возможности влияния на процесс самоорганизации, изучение характерных свойств образующихся структур.

Свойства квантовых точек

Применяя в качестве исходного исследуемого материала коллоидный водный раствор квантовых точек CdTe, считаем необходимым дать более подробное определение самому понятию квантовой точки.

Квантовые точки (КТ) – полупроводниковые нанокристаллы размером 2–10 нм, которые благодаря их уникальным оптическим, химическим, физическим свойствам являются перспективным объектом для применений в оптоэлектронике, электронике, фотонике, биологии и медицине.

Из-за малого размера подобные кристаллы обладают свойствами, отличными от «больших» полупроводниковых кристаллов. Пространственное ограничение движения носителей заряда приводит к квантово-размерному эффекту, выражающемуся в дискретной структуре электронных уровней, из-за чего КТ иногда называют «искусственными атомами» [3].

Благодаря эффекту квантового пространственного ограничения оптические параметры квантовых точек (спектры поглощения и люминесценции) зависят от размеров нанокристаллов. Поэтому, меняя их размеры и химический состав, можно получить спектры излучения квантовых точек в широкой области длин волн – от ближнего ультрафиолета до ближнего ИК диапазона. Спектр люминесценции КТ представляет собой относительно узкую полосу, положение максимума которой зависит от среднего размера КТ, а ширина определяется разбросом КТ по размерам. Обычно характерный разброс не превышает 5–10%, в результате чего эта полоса значительно уже, чем у большинства люминесцирующих органических молекул.

Способы получения полупроводниковых квантовых точек различны: они могут создаваться из планарных полупроводниковых гетероструктур с помощью литографии, с помощью молекулярно-лучевой эпитаксии и, наконец, могут получаться в результате высокотемпературного метало-органического химического синтеза.

Квантовые точки имеют большой коэффициент поглощения (до единиц $10^6 \text{ M}^{-1} \text{ cm}^{-1}$) в широкой полосе и высокий «квантовый выход» (до 80%). Таким образом, квантовые точки разного размера можно возбудить светом с одной и той же длиной волны, регистрируя люминесцентное излучение от них на разных длинах волн. Нанокристаллы обладают высокой химической и фотостабильностью.

Квантовые точки, получаемые в результате химического синтеза, как правило, не растворимы в воде, которая является естественной биологической средой. Однако существует ряд технологий, позволяющих делать КТ растворимыми в воде: создание молекулярной оболочки на поверхности КТ с молекулами, содержащими гидрофильные группы. Замена органической оболочки другими лигандами позволяет использовать КТ в качестве функциональных структур для различных целей, например, для присоединения их к биологическим макромолекулам. В частности, наши образцы CdTe КТ стабилизированы отрицательно заряженными молекулами меркаптоуксусной кислоты SHCH_2COOH (Thioglycolic acid, TGA), OH группа которой обеспечивает растворимость CdTe в воде.

Важными параметрами квантовых точек, знание которых необходимо при их использовании, являются: средний размер (D), коэффициент экстинкции (ε) и их концентрация (C_M) в используемом растворе. Все эти параметры можно определить в результате измерений спектров поглощения и люминесценции образцов растворов квантовых точек [4].

Для определения размера квантовых точек (D) различного состава обычно используются полуэмпирические выражения, в частности, для КТ на основе CdTe они имеют следующий вид:

$$D = (9.8127 \times 10^{-7}) \times \lambda^3 - (1.7147 \times 10^{-3}) \times \lambda^2 + (1.0064) \times \lambda - (194.84) , \quad (1)$$

где D – размер частиц (нм), λ – положение максимума длинноволновой полосы поглощения (нм). Используя полученные данные по размерам квантовых точек, можно вычислить молярный коэффициент экстинкции по следующим выражениям:

$$\text{CdTe: } \varepsilon = 10043 \times (D)^{2.12} ,$$

(2)

где ε – коэффициент экстинкции ($M^{-1}\cdot cm^{-1}$), D – размер частиц (нм). В свою очередь, концентрация в исследуемом растворе определяется по формуле:

$$C_M = \frac{A_{\max abs}}{\varepsilon \times l}, \quad (3)$$

где C_M – концентрация квантовых точек (M , где M – моль/литр = $6.3 \cdot 10^{23}$ штук/ cm^3), $A_{\max abs}$ – величина оптической плотности в максимуме длинноволновой полосы поглощения раствора квантовых точек, ε – коэффициент экстинкции ($M^{-1}\cdot cm^{-1}$), l – толщина слоя (см) (в нашем случае используются кюветы с $l = 1$ см).

Метод исследования

Среди различных строительных блоков нанометрового размера, используемых для создания упорядоченных структур применимых в оптоэлектронике, фотонике и производства биосенсоров, полупроводниковые нанокристаллы вызывают особый интерес. Это связано с тем, что основные свойства нанокристаллов, влияющие на процесс создания упорядоченных структур, такие как их размер, форма, защитная оболочка, стабилизирующие свойства (стабилизаторы) и заряд, могут контролироваться. Метод, используемый в работе для изучения данных процессов, условно называется «лаборатория в капле» [1]. Такой модельный подход позволяет изучать основные закономерности и движущие силы термодинамически индуцированных процессов самоорганизации НК на полярных поверхностях. В работе [1] было продемонстрировано, что при инкубации и контролируемом испарении водной капли раствора нанокристаллов могут быть получены поликристаллические дендриты с различной морфологией и размерами, зависящими от формы нанокристаллов, от точно подобранной концентрации и диапазона температур. Поэтому эта операция «лаборатория в капле» при контроле внешних параметров обеспечивает образование флуоресцентных структур выбранного размера и морфологии.

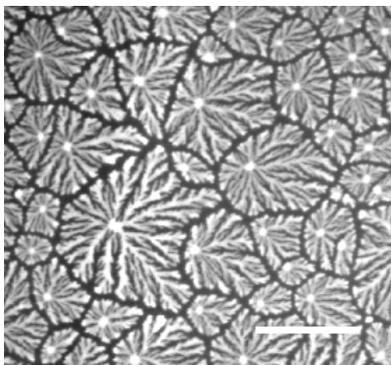


Рис. 1. Фотография дендритных структур, заполняющие 90% поверхности

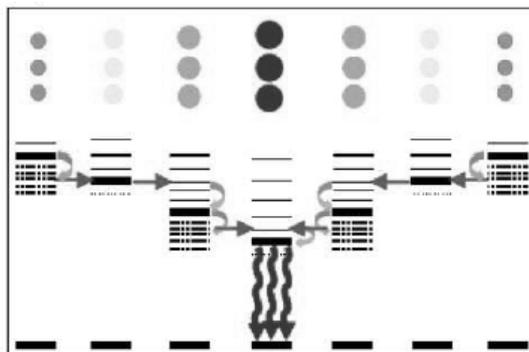


Рис. 2. Схематичная иллюстрация каскадного переноса поглощенной световой энергии от периферийных областей дендрита к его ядру

Контролируемая самоорганизация полупроводниковых НК позволяет создать упорядоченные дендритные структуры, заполняющие 90% поверхности светопоглощающим слоем (рис. 1), с морфологией, которая предоставляет возможности для каскадного переноса поглощенной световой энергии от периферийных областей дендрита к его ядру (рис. 2). Возможность каскадного переноса связана с тем, что дендритные структуры образованы НК, размер которых постепенно увеличивается от периферийных областей к центру дендрита. В то же время, контактирующие или близко расположенные нанокристаллы разного размера характеризуются высокоэффективным безызлучательным переносом энергии от частицы меньшего размера к частице большего размера.

Принципы образования дендритов

Поскольку целью данной работы является получение структур, называемых дендритами, следует сформулировать некоторые основные принципы их образования, разработанные для «классических» дендритных структур, выращенных из переохлажденных расплавов, которые, тем не менее, дают общее представление об образовании таких структур. Несмотря на особенности нашего метода, пересыщенный раствор является подобным переохлажденному расплаву своей метастабильностью и термодинамически неустойчивым состоянием и поэтому общие положения этих двух процессов схожи.

Дендриты (от греч. дендрон – дерево) – скелетные или ветвистые кристаллические структуры. При обсуждении вопросов образования структур полезно помнить, что:

- 1) рост дендритов идет в строго определенных направлениях;
- 2) ветвление происходит приблизительно через одинаковые промежутки, которые становятся все короче с повышением порядка ветвей;
- 3) дендритно затвердевает лишь небольшая доля расплава (раствора).

Из-за флуктуационного характера процесса кристаллизации и механизма роста граней граница кристаллов с переохлажденным раствором не может быть абсолютно гладкой, а характеризуется наличием некоторых выступов (бугорков) (рис. 3, а). Поскольку кристаллизация сопровождается выделением тепла, то исходное распределение температуры соответствует показанному на рис. 3, б. Это связано с тем, что вблизи поверхности расплава температура внешней среды будет больше, чем на расстоянии от нее, так как расплав отдает тепло внешней среде. Следовательно, у кончика бугорка температура ниже, а степень пересыщения выше, чем у основания. Поэтому бугорок удлинится и станет остроконечным. В то же самое время рост плоскости у его основания затормозится из-за выделения скрытой теплоты таким эллипсоидом. Начнут образовываться и другие остроконечные бугорки на расстояниях, определяемых радиусом зоны действия первого отростка. Таким образом, в конце концов вырастают ряды остроконечных отростков. Их боковой рост задерживается благодаря выделению скрытой теплоты плавления в соседних отростках, так что они растут преимущественно перпендикулярно поверхности раздела.

Возникновение ветвящихся отростков объясняется во многом сходными причинами. Первичный отросток, образовавшийся изложенным выше путем, находится в зоне температурной инверсии. Равномерный радиальный рост приблизительно цилиндрического первичного отростка должен быть неустойчивым, благодаря чему через промежутки, длина которых определяется количеством выделяющейся в ответвлениях скрытой теплоты плавления, станут возникать ряды новых отростков [5].

Периодичность возникновения ветвлений в дендритах металлов или льда заставляет предполагать, что объем кончика дендрита и распределение температуры вокруг последнего подвержены периодическим изменениям. Кончик дендрита растет до тех пор, пока не станет неустойчивым, после чего он делится на несколько отдельных кончиков, каждый из которых имеет меньший радиус кривизны. Всякий такой вновь обра-

зовавшийся кончик опять растет до тех пор, пока не обратится в неустойчивое состояние, и т.д.

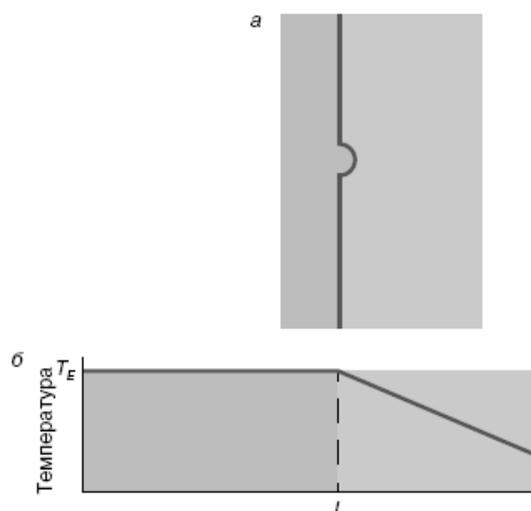


Рис. 3. Неустойчивость поверхности раздела в условиях температурной инверсии

Экспериментальная часть

Эксперименты проводились путем нанесения капли (8 мкл) коллоидного раствора CdTe квантовых точек с отрицательно заряженной поверхностью на стеклянную подложку, покрытую гидрофобным или гидрофильным слоем, и медленного высушивания этой капли. Обнаружено, что на поверхности подложки образуются дендритные структуры различной формы и размера. Образовавшиеся структуры анализировались с использованием метода фазового контраста и люминесценции на обычном (Jenaval contrast Zeiss) и люминесцентном (Люмам-ИЗ, ЛОМО) микроскопах. Экспериментально показано, что параметры дендритов определяются концентрацией КТ в растворе, температурой и скоростью высыхания капли. Скорость испарения воды (v_{ev}) капли раствора не должна превышать трансляционной скорости (v_t) и ориентационной скорости диффузии (v_{or}) КТ в капле для образования необходимых структур. В противном случае частицы, не имея необходимого времени для диффузного сближения при быстром высыхании коллоидного раствора, останутся прилипшими к поверхности подложки (этот метод используется для получения отдельных частиц на поверхности подложки). Для этого необходимо в процессе высыхания обеспечить постоянную температуру и зафиксировать локальное давление водяного пара. Концентрация раствора должна быть достаточно высокой, чтобы перенасыщенный раствор при высыхании заполнил все поле исследуемой поверхности, но не слишком большой, так как тогда будет образовываться многослойная структура, что существенно затруднит исследование новообразований. В настоящей работе в качестве начальной концентрации использовалась концентрация раствора КТ (4 мг/мл), используемая в работе [1], и в дальнейшем анализировались изменения получаемых структур при изменении концентрации. Был произведен перерасчет концентрации для CdTe с учетом отличающихся параметров нанокристаллов от представленных в статьях CdSe/ZnS.

В качестве подложки были взяты подготовленные предметные стекла для микроскопии, модифицированные положительно заряженными гидрофобным (silane) и гидрофильным (poly-L-lysine) покрытием, позволяющим избирательно менять вклад капиллярных сил в процесс самоорганизации.

На рис. 4–5 для иллюстрации представлены типичные спектры поглощения и люминесценции одного из исходных водных растворов КТ (образец №1). По полученным из измерений данным проводился расчет средних размеров КТ, концентрации и коэф-

фициента экстинкции. Для измерения спектров поглощения и люминесценции растворов КТ использовался спектрофлуориметр ФЛЮОРАТ-02-ПАНОРАМА.

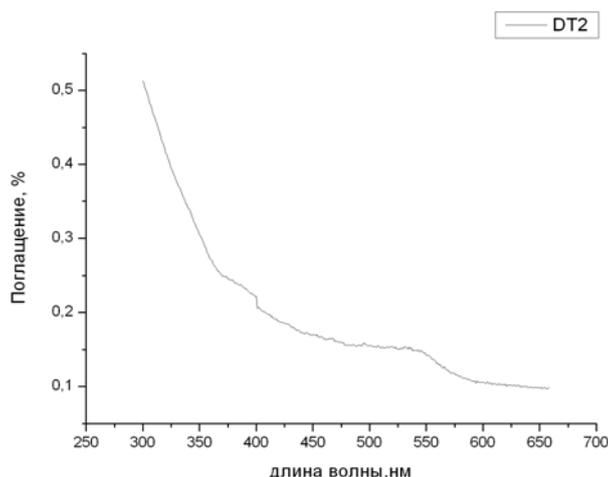


Рис. 4. Спектр поглощения водного раствора №1

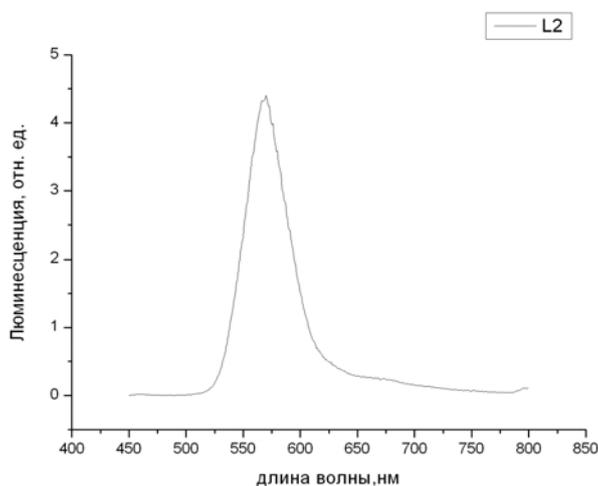


Рис. 5. Спектр люминесценции водного раствора №1

В таблице представлены параметры используемых исходных растворов КТ, рассчитанные из полученных графиков поглощения и люминесценции: концентрация (C_M), коэффициент экстинкции (ϵ) и средний размер (D).

№ образца	C, M	$\epsilon, \text{см}^{-1}M^{-1}$	$\lambda_{\text{maxabs}}, \text{нм}$	$\lambda_{\text{maxlum}}, \text{нм}$	$D, \text{нм}$
1	$1,50 \cdot 10^{-4}$	106732	534,5	569	3,05
2	$2,62 \cdot 10^{-4}$	103901	531,9	559,4	3,01
3	$2,29 \cdot 10^{-4}$	75377	509,6	543,6	2,59

Таблица. Основные параметры растворов КТ

Капля 8 мкл коллоидного раствора квантовых точек наносилась на стеклянную подложку и медленно высушивалась при различных внешних условиях. На рис. 6–11 представлены фазово-контрастные изображения (размер участка на всех снимках 230 мкм) дендритных структур с различной морфологией, полученные при разных внешних условиях (температуре, покрытии подложки) и концентрациях (значения приведены в подписях к рисункам). Эти результаты были получены в экспериментах с полностью закрытой чашкой Петри, когда скорость испарения раствора относительно невелика. При увеличении скорости испарения образование структур было затруднено.

Это подтверждает наше предположение о необходимости повышать локальное давление водяного пара, уменьшающее скорость испарения капли коллоидного раствора и, следовательно, увеличивающее время для диффузного сближения и взаимоориентации нанокристаллов.

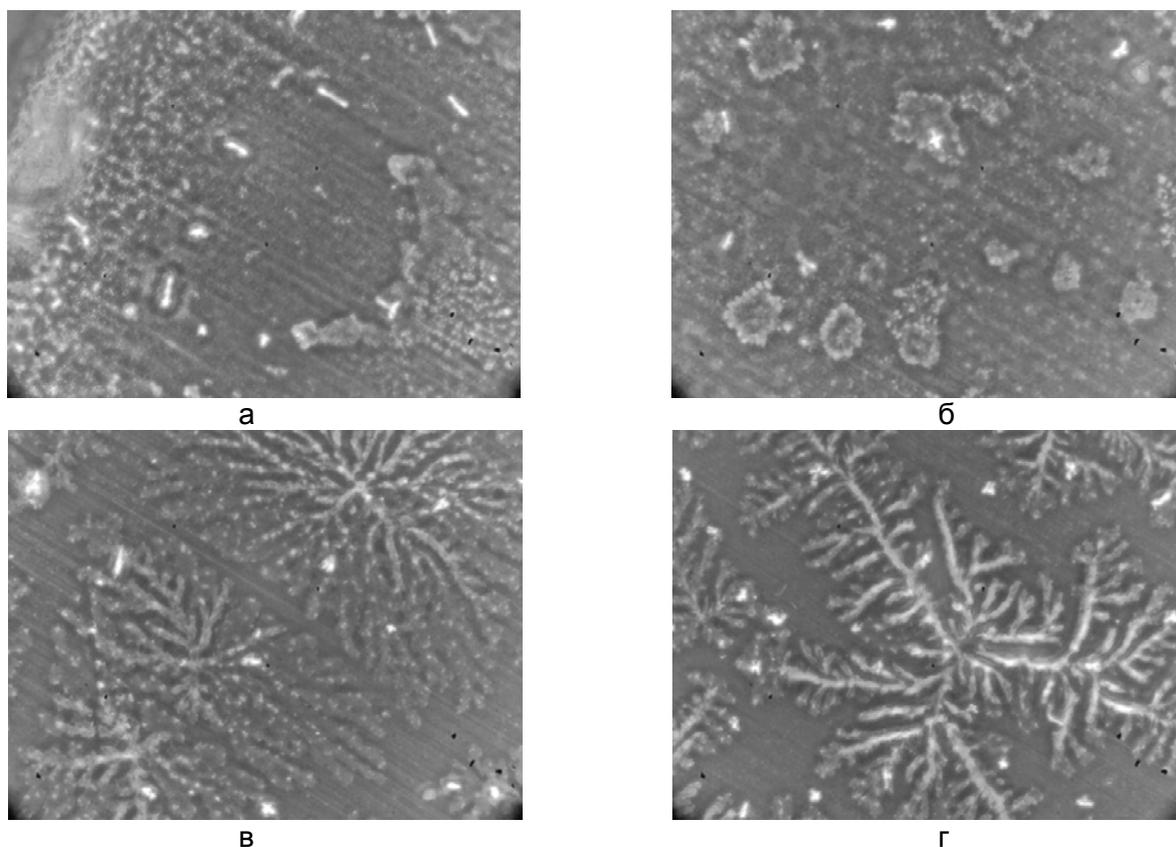


Рис. 6. Микрофотографии дендритных образований полученных при высыхании раствора КТ №2, разбавленного в 10 раз, на гидрофобной подложке, при температуре $\sim 10^\circ\text{C}$. Фотографии расположены в порядке удаления от края образца (а–г)

В серии снимков (рис. 6, а–г) изображен участок образца, начиная от края и следуя к центральной его части. По ним можно проследить изменения в образовании структур. В частности, это свидетельствует о том, что находящиеся в центре квантовые точки имели больше времени для диффузионного сближения и взаимоориентации

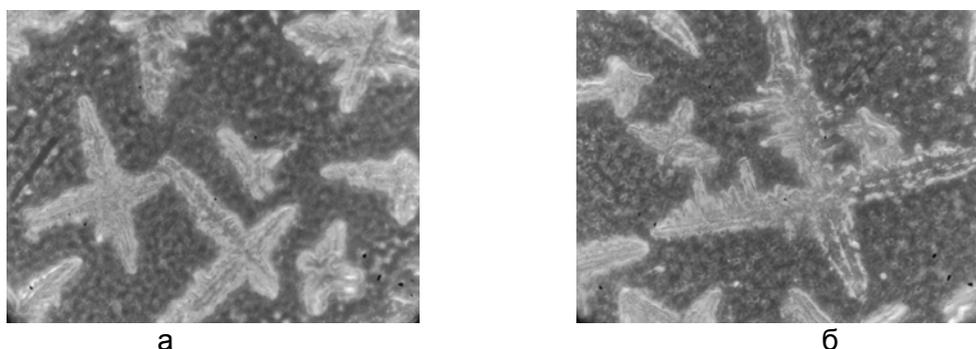


Рис. 7, а, б. Микрофотографии дендритных образований, полученных при высыхании на гидрофобной подложке раствора КТ №2, разбавленного в 50 раз, при температуре $\sim 10^\circ\text{C}$. Только на небольшом участке поля образовались ярко выраженные дендритные структуры в форме четких крестов с прямыми ответвлениями

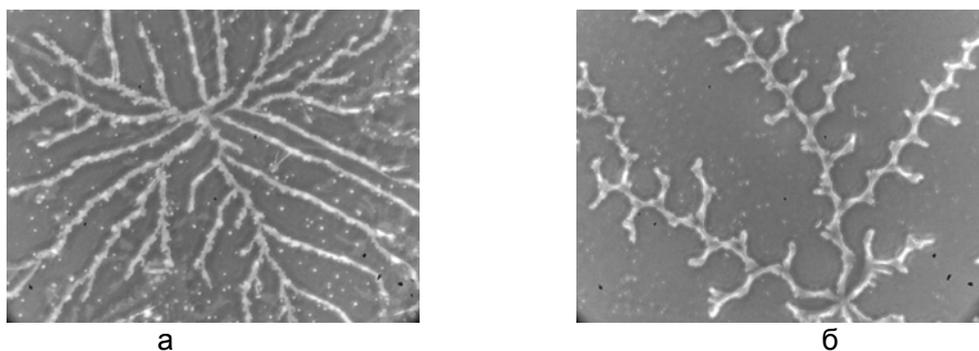


Рис. 8, а, б. Микрофотографии дендритных образований полученных при высыхании на гидрофобной подложке раствора КТ №2, разбавленного в 10 раз (а) и разбавленного в 50 раз (б), при температуре $\sim 22^{\circ}\text{C}$

На фотографиях несложно проследить различие размеров и морфологии полученных структур и их зависимость от концентрации. Также мы можем видеть некоторые проявления упорядоченности в полученных дендритах, особенно на рис. 8, б.

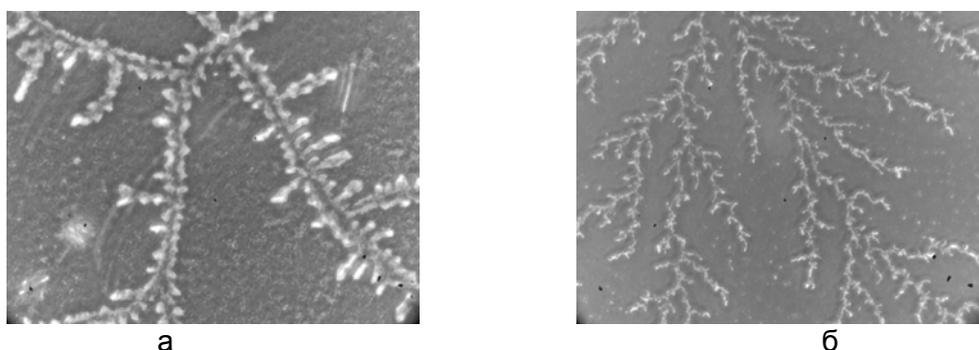


Рис. 9, а, б. Микрофотографии дендритных образований полученных при высыхании на гидрофильной подложке раствора КТ №2, разбавленного в 10 раз (а) и разбавленного в 50 раз (б), при температуре $\sim 22^{\circ}\text{C}$

Сравнение структур, образовавшихся на гидрофобной (рис. 8 и 10) и гидрофильной (рис. 9 и 11) подложках, показывает их явное отличие. Это показывает, что тип покрытия подложки влияет на характер морфологии структур при разных температурах и концентрациях.

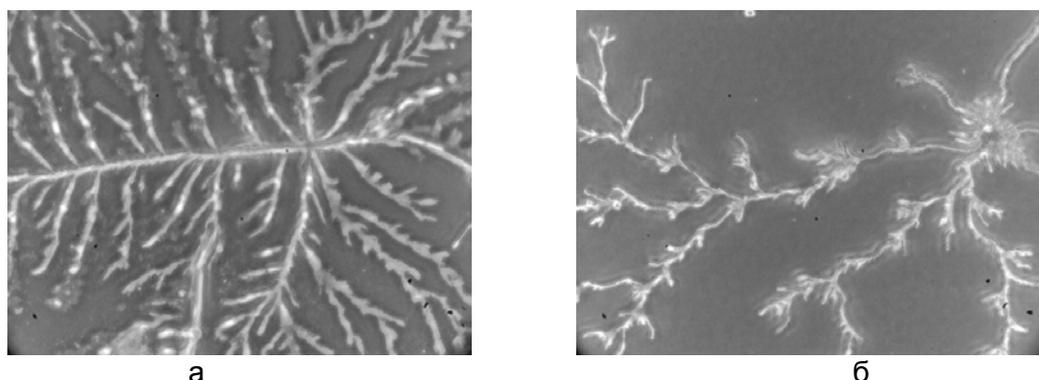


Рис. 10, а, б. Микрофотографии дендритных образований, полученных при высыхании на гидрофобной подложке раствора КТ №2, разбавленного в 10 раз (а) и разбавленного в 50 раз (б), при температуре $\sim 4^{\circ}\text{C}$

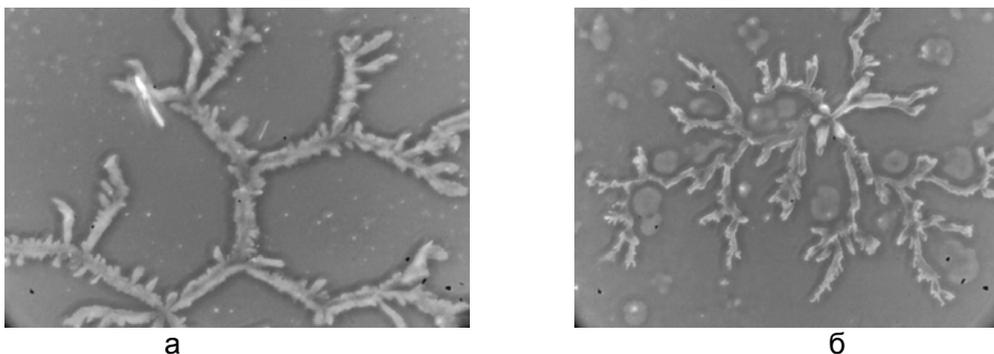


Рис. 11, а, б. Микрофотографии дендритных образований, полученных при высыхании на гидрофильной подложке раствора КТ №2, разбавленный в 10 раз (а) разбавленный в 50 раз (б), при температуре $\sim 4^{\circ}\text{C}$

Заключение

Были проведены исследования заряженных водорастворимых нанокристаллов CdTe на наличие у них способности к самоорганизации в процессе высыхания раствора КТ на модифицированных (гидрофильных и гидрофобных) поверхностях твердого тела. В результате этого были получены дендритные структуры различной морфологии и размеров, во многом близкие к полученным ранее структурам на основе квантовых точек CdSe/ZnS. В экспериментах также прослеживаются аналогичные зависимости морфологии полученных структур от воздействия внешних условий самоорганизации. Сделан вывод о том, что основные закономерности самоорганизации являются общими для КТ различных материалов.

Литература

1. Alyona Sukhanova, Yuri Volkov, Andrey L Rogach, Alexander V Baranov, Andrei S Susha, Dmitriy Klinov, Vladimir Oleinikov, Jacques H M Cohen and Igor Nabiev. Lab-in-a-drop: controlled self-assembly of CdSe/ZnS quantum dots and quantum rods into polycrystalline nanostructures with desired optical properties // *Nanotechnology*. – 2007. – 18. – 185602.
2. Alyona Sukhanova, Alexander V Baranov, Dmitriy Klinov, Vladimir Oleinikov, Kevin Berwick, Jacques H M Cohen, Michel Pluot and Igor Nabiev. Self-assembly of charged microclusters of CdSe/ZnS core/shell nanodots and nanorods into hierarchically ordered colloidal arrays // *Nanotechnology*. – 2006. – 17. – 4223.
3. Федоров А.В., Баранов А.В. Оптика квантовых точек / Оптика наноструктур. Под ред. А.В. Федорова. – СПб: Недра, 2005. – С. 181–274.
4. W. W. Yu, L. Qu, W. Guo, and X. Peng. Experimental Determination of the Extinction Coefficient of CdTe, CdSe, and CdS Nanocrystals // *Chem. Mater.* 2003. – 15. – 2854.
5. Третьяков Ю.Д.. Дендриты, фракталы и материалы. // *Соросовский образовательный журнал*. – 1998. – 11. – 96.
6. J. Dutta and H. Hofmann. Self-Organization of Colloidal Nanoparticles // *Switzerland Encyclopedia of Nanoscience and Nanotechnology*. – 2003. – V. 10. – 1.
7. D.V. Talapin, E.V. Shevchenko, C. B. Murray, A. Kornowski, S. Forster, and H. Weller. CdSe and CdSe/CdS Nanorod Solids // *J. Am. Chem. Soc.* – 2004. – 126. – 12984.

ЭЛЕМЕНТАРНЫЕ ТЕЧЕНИЯ В НАНОГИДРОДИНАМИКЕ

А.И. Свитенков, В.В. Лесничий

Научный руководитель – к.ф.-м.н., доцент С.А. Чивилихин

Рассматривается течение Пуазейля с учетом особенностей наногидродинамики. Основное внимание уделено соизмеримости длины свободного пробега частицы с характерными линейными размерами системы (диаметр трубки). В работе производится написание новых, отличных от классических, уравнений течения и проверка полученного результата методом молекулярной динамики.

Введение

В настоящее время развитие науки, занимающейся рассмотрением процессов протекания жидкости (газа) в системах с характерными линейными размерами порядка от 1 мкм до 1 нм и получившей название «наногидродинамика», сталкивается с определенными трудностями. Последние заключаются в том, что уравнения наногидродинамики еще не записаны, а инструментарий исследований практически состоит из одного лишь метода моделирования молекулярной динамики. Несмотря на высокие требования к производительности вычислительных платформ, данный метод получил широкое распространение и используется достаточно успешно [1, 2]. Так, например, относительно недавно целый поток работ различного уровня был направлен на изучение явления проскальзывания в пристеночном слое, в результате чего удалось сформировать отличные от классических граничные условия, учитывающие данную особенность [2, 3]. Однако результативность исследований значительно снижается за счет необычайно широкой области рассматриваемых эффектов. Выделяют следующие особенности микро- и наногидродинамики:

- размеры канала сравнимы с размерами молекул, движущихся по каналам;
- флуктуации плотности могут быть велики, в отличие от макротечений;
- транспортные свойства (вязкость, диффузия, теплопроводность) могут содержать размерные факторы (как в турбулентности);
- взаимодействие нанопотока со стенкой может быть определяющим фактором;
- точной формы граничных условий нет;
- приближение сплошной среды может нарушаться;
- в микро- и нанопотоках наблюдаются явления, не существующие в макроскопической гидродинамике.

Таким образом, нарушаются практически все модельные приближения, лежащие в основе классической гидродинамики [4]. Но до сих пор остается неясным, какие из них действительно носят принципиальный характер, а нарушение каких не скажется на формировании окончательных законов движения. В сложившейся ситуации наиболее логичной оказалась бы попытка ответить на этот вопрос, и, таким образом, сузить круг рассматриваемых проблем. Однако, пожалуй, другого способа исключения принципиальных эффектов и выделения основополагающих, кроме как перебор и последовательная проверка, нет.

В данной работе проводилось рассмотрение механизма вязкого переноса импульса, который лежит в основе классических уравнений Навье-Стокса, в рамках наногидродинамики. Рассматриваемое отличие от классического механизма состояло в соизмеримости длины свободного пробега частицы с характерными линейными размерами системы, т.е. рассматривалось течение разреженных сред. Полученные на основе данной гипотезы уравнения проверялись методами молекулярной динамики.

Постановка задачи

Рассмотрим подробнее задачи, в условиях которых производилось численное моделирование динамики жидкости, а также наметим план, в соответствии с которым проводи-

лось описываемое исследование. Как было заявлено выше, основное внимание уделялось особенностям механизма переноса импульса, и их влиянию на конечный результат.

Для максимального упрощения расчетов рассматривалась задача о течении жидкости через трубку – так называемое течение Пуазейля. Диаметр трубки варьировался от 1 до 10 нм.

Результатом моделирования явилось построение фронта скоростей, а также вычисление таких величин, как средняя длина свободного пробега частицы, плотность, средняя скорость движения частиц, необходимых для выполнения теоретических расчетов. Результатом последних, в свою очередь, также явилось построение фронта скоростей.

Проверка предлагаемого формализма заключалась в сопоставлении теоретического результата и результата, полученного в ходе моделирования.

Уравнение переноса импульса для центральной части трубки

Проведем рассмотрение процесса переноса импульса через границу между слоями жидкости аналогично классическому, однако учтем факт соизмеримости длины свободного пробега частицы и размеров самой системы.

Согласно модельным представлениям [5], количество импульса, переносимого через фрагмент границы слоя S в единицу времени, в направлении рассматриваемого слоя, равно:

$$\sigma = \frac{1}{6} n \cdot m \cdot \langle v \rangle (u_1 - u_0) \cdot S, \quad (1)$$

где $\langle v \rangle$ – средняя скорость теплового движения молекул (предполагается одинаковой для рассматриваемых слоев), u_1 и u_0 – скорости упорядоченного движения граничащего и рассматриваемого слоя, соответственно.

Выделим уровень с координатой x и скоростью $u(x)$ (рис. 1). В процесс переноса импульса через уровень будут участвовать слои, отстоящие от него на λ : $x+\lambda$, $x-\lambda$, где λ – средняя длина свободного пробега частицы. Подставим в (1):

$$\sigma = \frac{1}{6} n \cdot m \cdot \langle v \rangle (u(x - \lambda) - u(x + \lambda)) \cdot S. \quad (2)$$

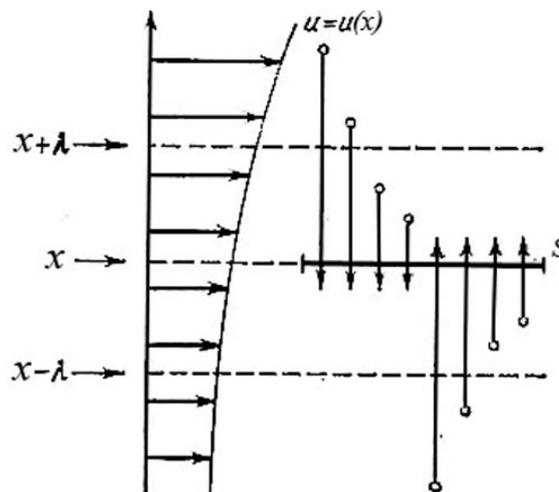


Рис. 1. Механизм переноса импульса

Для нахождения значений скорости в указанных точках разложим $u(x)$ в степенной ряд. Однако, вспоминая вышесказанное о длине свободного пробега, проведем разложение до 3-й степени (2-я степень разложения дает ноль, при формировании разности в (2)), в отличие от классического случая, когда учитывается только линейный член. Составим разность:

$$[u(x - \lambda) - u(x + \lambda)] = -2 \cdot u'(x_0)(\lambda) - \frac{1}{3} u'''(x_0)(\lambda)^3. \quad (3)$$

Подставим полученное выражение в (1):

$$\sigma = \frac{1}{6} n \cdot m \cdot \langle v \rangle [-2 \cdot u'(x_0)(\lambda) - \frac{1}{3} u'''(x_0)(\lambda)^3] \cdot S. \quad (4)$$

Переобозначим $\mu = \frac{1}{3} n \cdot m \cdot \langle v \rangle$ и подставим полученное выражение в уравнение течения жидкости по трубе [4]:

$$-\frac{\partial \sigma}{\partial x} = \frac{\partial P}{\partial y}. \quad (5)$$

Заметим, что левая и правая части здесь зависят о разных переменных, откуда можно сделать вывод, что правая часть – константа. Подставляя выражения для σ , получим:

$$\mu(\lambda \cdot \frac{d^2}{dx^2} u(x) + \frac{1}{6} \cdot \lambda^3 \cdot \frac{d^4}{dx^4} u(x)) = k. \quad (6)$$

Интегрирование уравнения (6) дает:

$$u(x) = -\frac{1}{6} \cdot C2 \cdot \lambda^2 \cdot \sin\left(\frac{\sqrt{3} \cdot x}{\lambda}\right) - \frac{1}{3} \cdot C1 \cdot \lambda^2 \cdot \cos\left(\frac{\sqrt{3} \cdot x}{\lambda}\right) - \frac{1}{2} \frac{k \cdot x^2}{\mu \cdot \lambda} + C3x + C4.$$

Из соображений симметрии полагаем $C2=0$ и $C3=0$; окончательно уравнение принимает вид:

$$u(x) = -\frac{1}{6} \cdot C1 \cdot \lambda^2 \cdot \cos\left(\frac{\sqrt{3} \cdot x}{\lambda}\right) - \frac{1}{2} \frac{k \cdot x^2}{\mu \cdot \lambda} + C4. \quad (7)$$

Хорошо видно, что при переходе в область малых значений λ уравнение стремится к классическому виду, описывающему параболический фронт скоростей $u(x)$, что в определенной мере является критерием истинности полученного результата.

Однако уравнение (7) содержит неизвестную константу $C1$, значение которой не может быть определено из классических граничных условий или граничных условий проскальзывания; кроме того, первое слагаемое не подходит по размерности, которая, очевидно, должна составлять [м/с]. Для ликвидации этих недостатков необходимо рассмотреть процесс переноса импульса в пристеночном слое трубки.

Уравнение переноса импульса для пристеночного слоя трубки

Механизм переноса импульса в пристеночном слое приобретает несколько иной вид, нежели описанный выше. Для рассмотрения введем систему координат, совместив ось ОУ со стенкой трубки. Далее нас будет интересовать область $x < (\lambda - r)$. Ее особенность заключается в том, что в механизме переноса импульса через рассматриваемый уровень x участвует вещество, локализованное $x+\lambda$, $x+\delta$, где $\delta = \lambda - 2x$. Данный механизм поясняется на рис. 2. Здесь для простоты расчетов мы делаем довольно серьезное допущение, условившись считать, что при взаимодействии со стенкой продольная и поперечная составляющие импульса сохраняются. Очевидно, это не так, и впоследствии от данного допущения следовало бы отказаться, однако в рамках проводимого исследования указанная условность значительно не повлияет на общий результат и сформулированные выводы.

Итак, проведем разложение $u(x)$ в степенной ряд до линейного члена. Суммарный перенос импульса запишется как

$$u(\lambda - x) + u(x + \lambda) = 2 \cdot u(x) + 2 \cdot u'(x)(\lambda - 2x). \quad (8)$$

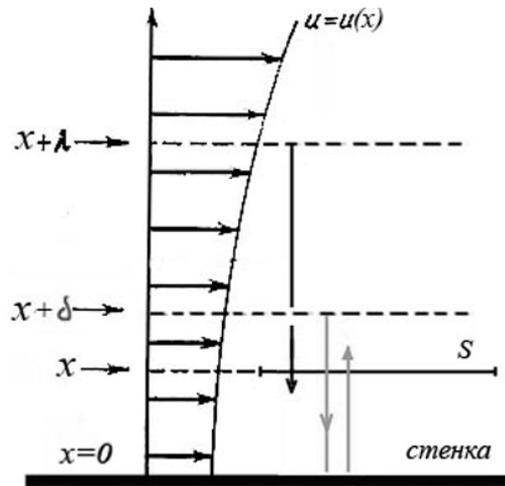


Рис. 2. Особенности переноса импульса в пристеночном слое $x < (\lambda - r)$

Так как поток импульса через единицу сечения не может увеличиваться за счет потока вещества в силу постоянства последнего при фиксированных плотности и температуре, то полученную сумму следует разделить на 2. Итоговое выражение для переноса импульса

$$\sigma = \frac{1}{6} n \cdot m \cdot \langle v \rangle [u(x_0) + u'(x_0)(\lambda - x_0)] \cdot dS. \quad (9)$$

Подставляя в выражение (5), получаем уравнение, описывающее течение жидкости в пристеночном слое:

$$-\mu \cdot (\lambda - x) \left(\frac{d^2}{dx^2} \cdot u(x) \right) = k. \quad (10)$$

Интегрирование дает:

$$u_{кр}(x) = -\frac{k}{\mu} \cdot ((\lambda - x) \ln(\lambda - x) - (\lambda - x)) + C5 \cdot x + C6. \quad (11)$$

Обратим внимание на константу C5. Если рассматривать процесс переноса импульса только при участии уровня $x + \lambda$, не принимая во внимание роль уровня $x + \delta$, то перенос импульса окажется пропорционален первой производной $u(x)$. Подобный механизм уже был описан в ряде работ [6], а результат рассмотрения был сформулирован как граничное условие проскальзывания.

$$U|_S = l \frac{\partial U}{\partial n} \Big|_S.$$

Здесь S – граница, в нашем случае – стенка трубки, имеющая координату H, а l – коэффициент, так называемая длина проскальзывания. Тогда (для простоты вычислений производная берется от параболического профиля скорости [6]) $C5 = k/\mu$. Окончательный вид для (11):

$$u_{кр}(x) = -\frac{k}{\mu} \cdot ((\lambda - x) \ln(\lambda - x) - (\lambda - x)) + l \frac{k}{\mu} \cdot x + C6. \quad (12)$$

Сшивка

Для нахождения оставшихся констант выполним сшивку функций $u(x)$ и $u_{кр}(x)$ в точке $x = \lambda - r$, где r – радиус частицы, для системы координат, значения x в которой отсчитывается от стенки (в ней задано $u_{кр}(x)$), соответствующей точке $x = H - (\lambda - r)$ в системе координат, где x отсчитывается от центра трубки. Здесь H – радиус трубки. Условие сшивки:

$$\begin{aligned} u_{кр}(x)|_{x=\lambda-r} &= u(x)|_{x=H-(\lambda-r)} \\ u'_{кр}(x)|_{x=\lambda-r} &= u'(x)|_{x=H-(\lambda-r)} \end{aligned} \quad (13)$$

Наибольший интерес для нас представляет константа $C1$, которое находится из условий равенства производных. Вычислим ее значение.

$$C1 = -\frac{2 \cdot \sqrt{3} \cdot k}{\mu} \cdot \frac{(\lambda \cdot \ln(\lambda - x) + \lambda + r - H + l \cdot H)}{\lambda^2 \sin\left(\frac{\sqrt{3} \cdot (\lambda - r - H)}{\lambda}\right)}. \quad (14)$$

На этом подготовительные перед моделированием операции можно считать законченными. Значение $C4$ найдется численно, исходя из значения скорости в центре трубки, полученной в результате моделирования. При аппроксимации результатов симуляции будет использовано только уравнение (7). Уравнения для пристеночного слоя (12) при аппроксимации не учитываются.

Моделирование молекулярной динамики

Моделирование движения частиц осуществляется путем расчета взаимодействия каждой молекулы (атома) с окружающими ее в соответствии со вторым законом Ньютона:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \sum_j \mathbf{F}_{ij},$$

где m_i – масса i -й частицы, \mathbf{r}_i – координата i -й частицы, \mathbf{F}_{ij} – сила взаимодействия i -й и j -й частицы. Сила взаимодействия находится в соответствии с выражением $\mathbf{F}_{ij} = -\nabla U_{ij}$.

Для описания межмолекулярного взаимодействия использовался потенциал Леннарда-Джонса (7–13) [7]:

$$U(r) = \varepsilon \left(\left(\frac{\sigma}{r} \right)^{12} - 2 \left(\frac{\sigma}{r} \right)^6 \right),$$

где ε – глубина потенциальной ямы, σ – равновесное расстояние, r – расстояние между взаимодействующими молекулами (рис. 3).

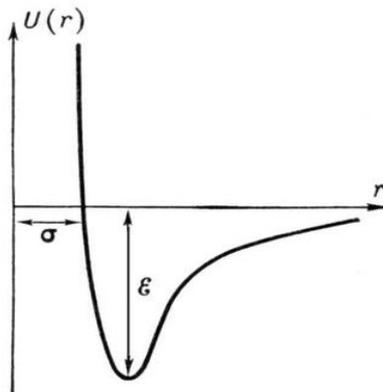


Рис. 3. Потенциал Леннарда-Джонса

Тогда сила взаимодействия между двумя wybranными молекулами найдется из формулы:

$$F(r) = 12 \frac{\varepsilon}{\sigma} \left(\left(\frac{\sigma}{r} \right)^{13} - \left(\frac{\sigma}{r} \right)^7 \right).$$

При расчете учитывалась симметричность матрицы взаимодействия F_{ij} , что позволило рассчитывать силы взаимодействия только для половины всех частиц. Таким

образом, производительность возросла почти в 2 раза. Также, для увеличения общей производительности системы, учитывалось взаимодействие только близлежащих молекул, т.е. допускалось увеличение расстояния взаимодействия по сравнению с равновесным в два раза [8].

Результаты моделирования. Аппроксимация

Моделирование течения производилось для трубок диаметром 3, 4 и 6 нм. При этом результатом работы программы являлось нахождение профиля скоростей. Полученный набор значений аппроксимировался посредством уравнения (7) и при учете (14). Подгоночными коэффициентами служили:

- l – длина проскальзывания;
- λ – длина свободного пробега частицы;
- k/μ – отношение градиента давления к «аналогу» (не учитывается λ) вязкости;
- $C4$ – скорость течения в центре трубки.

Наглядные результаты аппроксимации ($H = 3$ нм) представлены на рис. 4. На графике черными точками нанесены значения профиля скоростей, полученные в результате моделирования. Красная линия представляет собой график уравнения (7) с наилучшим образом подобранными «подгоночными» коэффициентами.

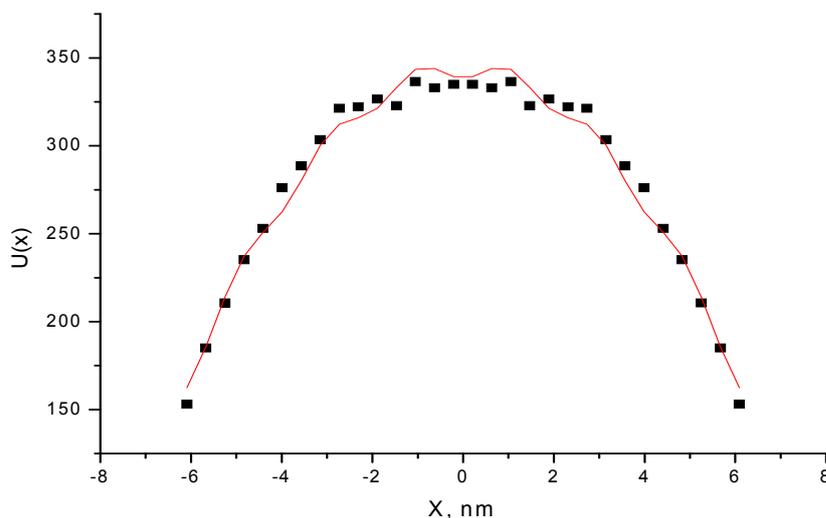


Рис. 4. Аппроксимация профиля скоростей, полученного в результате моделирования

Однако, пожалуй, субъективно кажущееся близким расположение на графике аппроксимационной кривой и экспериментальных данных не может составить достаточный критерий истинности или хотя бы пригодности предлагаемого формализма. Поэтому в ходе моделирования также измерялась средняя длина свободного пробега частицы. Полученное значение сравнивалось со значениями λ , найденными в результате аппроксимации. Очевидно, что их примерное равенство является безусловным подтверждением пригодности предлагаемых уравнений.

Совокупные результаты моделирования и аппроксимации представлены в таблице. Видно, что длина свободного пробега не всегда соответствует измеренной, величина достоверности аппроксимации R^2 не постоянна велика. А именно, достоверность оказывается намного ниже (иногда даже ниже, чем при аппроксимации классическим уравнением параболы), как было выяснено в ходе исследований, для значений $CI > 0$. Причина данного явления еще подлежит изучению, а пока, таким образом, мы определили область применимости описываемой модели.

Радиус трубки H, нм	Измеренное значение длины свободного пробега λ , нм	Теоретическое значение длины свободного пробега λ , нм	Величина дос- товерности ап- проксимации R^2	Величина дос- товерности ап- проксимации R^2 для парабо- лического про- филя
1,5	0,301	0,291	0,9937	0,9957
1,5	0,532	0,393	0,9930	0,9631
1,5	0,480	0,498	0,9991	0,9673
2	0,412	0,400	0,9732	0,9862
2	0,553	0,543	0,9912	0,9667
2	0,637	0,642	0,9915	0,9731
3	0,503	0,517	0,9850	0,9831
3	0,657	0,669	0,9843	0,9820

Таблица. Результаты моделирования и аппроксимации

Заключение

В ходе работы было оценено влияние соизмеримости длины свободного пробега с линейными размерами системы. Уравнения профиля скоростей, написанные при учете этой особенности, имеют вид, отличный от классического, и, как было выяснено в ходе моделирования, позволяют значительно лучше описывать явление. Однако указанное достоинство предлагаемого аппарата проявляется не всегда, а, как показало исследование, лишь для некоторой (вполне определенной) области значений длин пробега, что является несомненным недостатком. Причины его, возможно, кроются в чрезмерно упрощенных представлениях о процессах, происходящих в пристеночном слое. Модельные приближения, используемые в этой области, безусловно, являются самым тонким местом в работе и в первую очередь требуют устранения в ходе развития темы.

Литература

1. Gaspard P. Out-of-equilibrium nanosystems – Center for Nonlinear Phenomena and Complex Systems, Universit'e Libre de Bruxelles, Belgium, 2005.
2. Shin-Pon Ju A molecular dynamics simulation of the adsorption of water molecules surrounding an Au nanoparticle // The Journal Of Chemical Physics. – 2005. – 122. – 094718.
3. Tiezheng Qian, Xiao-Ping Wang, Hydrodynamic slip boundary condition at chemically patterned surfaces: A continuum deduction from molecular dynamics. – Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, 2005.
4. Ландау Л.Д., Лифшиц Е.М. Теоретическая физика в 10 томах. Том 6. Гидродинамика – М.: Физматлит, 2004. – 735 с.
5. М. Н. Коган, Динамика разряженного газа. – М.: Наука, 1967. – 440 с.
6. Georg Karniadakis, Ali Beskok, Narayan Aluru, Microflows and Nanoflows. – New York, NY : Springer, 2005. – 808 с.
7. Itsuo Hanasaki, Akihiro Nakatani, Hiroshi Kitagawa Molecular dynamics study of Ar flow and He flow inside carbon nanotube junction as a molecular nozzle and diffuser // Science and Technology of Advanced Materials. – 2004. –5. – 107–113.
8. Ciccotti G., Hoover W. G., Molecular-dynamics simulation of Statistical-Mechanical Systems. – Varenna, Italy, 1985. – 622 с.

ОПТИЧЕСКИЕ МЕТОДЫ ОПРЕДЕЛЕНИЯ РАЗМЕРОВ И ФОРМ НАНОЧАСТИЦ

А.Е. Логунов, А.В. Папко

Научный руководитель – д.ф.-м.н., ст.н.с. Т.А. Варганиян

В настоящей работе предложен новый экспериментальный метод определения пространственной анизотропии наноструктур. Метод основан на вращении вектора поляризации сфокусированного лазерного излучения при его сканировании по поверхности исследуемого образца и совместим с условиями высокого вакуума.

Введение

Исследование структуры островковых пленок сопряжено с рядом сложных технических проблем. Разработанные к настоящему времени методы их исследования весьма трудоемки и сопровождаются либо уничтожением объекта исследования (электронная микроскопия), либо требуют непосредственного объединения напылительной камеры и измерительного оборудования (атомно-силовая микроскопия). Кроме того, оба эти высокоразрешающих метода дают информацию лишь о малых площадях пленки (порядка 1 мкм^2), а полученные данные требуют дополнительной обработки. Для решения указанных проблем был предложен и реализован оптический метод бесконтактного определения параметров структуры пленок, названный нами флуктуационно-поляризационной микроскопией.

Возможность определения таких характеристик островковой металлической пленки, как средняя плотность и средний размер наночастиц, была продемонстрирована ранее в работе [1]. Эти характеристики были получены за счет измерений флуктуаций интенсивности рассеянного и поглощенного света при сканировании сфокусированного луча по пленке. Флуктуации определялись двумя характеристиками: числом островков, попадающих в луч, и их сечениями рассеяния и поглощения – величинами, определяемыми размерами островков. Для случая расположения островков на поверхности подложки случайным образом была продемонстрирована возможность определить эти параметры, а также распределение островков по размерам за счет измерений флуктуаций интенсивностей отраженного и поглощенного света, а также их коррелятора.

Суть настоящей работы – усовершенствование метода флуктуационной спектроскопии [1] за счет вращении вектора поляризации сфокусированного лазерного излучения при его одновременном сканировании по поверхности исследуемого образца. Флуктуации интенсивности прошедшего света за счет изменения поляризации излучения будут определяться теперь также и степенью пространственной анизотропии частиц, приводящей к неодинаковому их поглощению при изменении направления линейной поляризации.

Поляризационная спектроскопия

Пусть в фокальном пятне линейно поляризованного лазерного излучения с напряженностью поля, равной E , и частотой вращения поляризации ω находятся $N \gg 1$ частиц. Поглощение света j -той частицей (рис. 1), сечения поглощения которой в двух взаимоперпендикулярных направлениях равны σ_1 и σ_2 , определяется выражением:

$$a_j = \sigma_1 E_1^2 + \sigma_2 E_2^2 = E^2 [\sigma_1 \sin^2(\alpha_j - \varphi) + \sigma_2 \cos^2(\alpha_j - \varphi)] = I_j [o_j + \sigma_j \cos 2(\alpha_j - \varphi)],$$

где $\varphi = \omega t$ – угол поворота поляризации, $o_j = (\sigma_1 + \sigma_2)$ и $\sigma_j = \sigma_1 - \sigma_2$. Поглощение же всеми частицами равно

$$A(\varphi) = \sum_j I_j [o_j + \sigma_j \cos 2(\alpha_j - \varphi)] = A_0 + A_2 e^{i2\varphi} + \text{к.с.}$$

Анизотропная составляющая поглощения осциллирует на частоте 2ω , а изотропная постоянна. При этом их амплитуды несоизмеримы. Среднее значение изотропного вклада $\sim N$ и отклонения от него $\sim \pm N^{1/2}$. Среднее значение анизотропного вклада равняется нулю. Среднее отклонение от него $\sim \pm N^{1/2}$, что и характеризует величину A_2 .

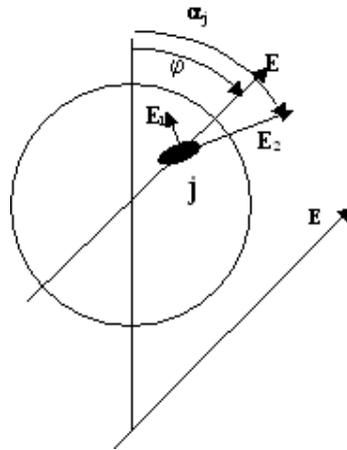


Рис. 1. Схематическое изображение наночастицы на поверхности подложки

Если бы число частиц в пятне оставалось неизменным, то вариации их ориентации в пятне вызывали бы 100%-ные изменения сигнала на частоте 2ω . Изменение числа частиц в пятне меняло бы амплитуду этого сигнала на величину $(N \pm N^{1/2})^{1/2} - (\pm)N^{1/2} = \pm 1/2 \ll N^{1/2}$ самой амплитуды. Поэтому в анизотропной составляющей можно не учитывать вклад флуктуаций числа частиц, попадающих в пятно и поглощающих излучение.

Если в неподвижном круглом пятне вращать поляризацию с угловой скоростью ω , то поглощение будет модулировано на частоте 2ω . И если при этом центр пятна движется по образцу со скоростью v , то вместе с этим будет проявляться модуляция поглощения на частотах порядка v/d (где d – диаметр пятна), обусловленная движением пятна и изменением в нем числа островков. Масштаб этих флуктуаций будет $\sim N^{1/2}$, т.е. таким же, как и на частоте 2ω .

Реализация предложенного принципа двойной модуляции поглощения позволяет определить сразу относительную среднюю асимметрию частиц островков по отношению интенсивности сигнала на частоте 2ω к интенсивности сигнала на частоте вращения луча по поверхности кюветы. Так как интенсивности обоих сигналов $\sim N^{1/2}$, то в их отношении исчезает зависимость от N , и остаются лишь параметры, характеризующие изотропную и анизотропную составляющие поглощения.

При изменении положения пятна света на пленке в нем оказываются разные частицы, и поэтому поглощение света флуктуирует. Выбором соответствующих частот вращения удастся разделить флуктуации поглощения разных гармоник модуляции. Флуктуации поглощения на частотах, связанных с движением луча по пленке, порождаются преимущественно флуктуациями изотропного поглощения, характеризуемого величиной $B_0 = [\langle (A_0 - \langle A_0 \rangle)^2 \rangle]^{1/2}$.

Флуктуации поглощения на гармониках модуляции вращения вектора поляризации лазерного излучения обусловлены анизотропией частиц и характеризуются величиной $B_2 = \langle A_2^2 \rangle^{1/2}$.

В предположении отсутствия корреляции между характеристиками разных частиц и их числом в пятне вычисление первой дает $B_0 = [N \langle \sigma^2 \rangle]^{1/2}$, и для второй получается $B_2 = [N \langle \sigma^2 \rangle]^{1/2}$, где N – среднее число поглощающих частиц. Отношение

$$B_0 / B_2 = [\langle \sigma^2 \rangle / \langle \sigma^2 \rangle]^{1/2} = [\langle (\sigma_1 + \sigma_2)^2 \rangle / \langle (\sigma_1 - \sigma_2)^2 \rangle]^{1/2}$$

не зависит от N . После преобразований данное выражение можно упростить и получить, что $B_0 / B_2 = (1 + \langle \sigma_1 / \sigma_2 \rangle) / (1 - \langle \sigma_1 / \sigma_2 \rangle)$.

В эксперименте измерялись интенсивности флуктуирующих сигналов гармоник, прошедших поглотитель. Эта величина пропорциональна интенсивности I_k , падающей на поглотитель k -той гармоники. Поэтому измеряемой величиной является $G = I_0 B_0 / I_2 B_2$, из которой определяется асимметрия поглощения частиц

$$\langle \sigma_1 / \sigma_2 \rangle = (G - 1) / (G + 1).$$

Эксперимент

В эксперименте использовались непрерывные одночастотные полупроводниковые лазеры мощностью 30 мВт с длинами волн 650 нм, 532 нм и 980 нм. Поляризация света изменялась нами за счет эффекта Погкельса. Лазерное излучение фокусировалось на поверхность островковой пленки микрообъективом с фокусным расстоянием 13,9 мм и числовой апертурой 0,3. Размер пятна на пленке составлял 10 мкм.

Островковые пленки создавались в запаянных цилиндрических вакуумированных кюветах посредством конденсации на холодном сапфировом окне кюветы паров натрия, полученных нагреванием капли металла на стенке кюветы.

Структурные характеристики пленок, полученных указанным образом, определялись ранее методом флукуационной спектроскопии [1]. В пленках с коэффициентом экстинкции < 0.3 получались поверхностные концентрации наночастиц $\sim 10^{10} - 10^{12} \text{ см}^{-2}$ и средними размерами 60–80 нм.

Типичный спектр экстинкции островковой пленки натрия на сапфире состоит из двух полос (рис. 2). Такой вид спектра экстинкции связан со специфической формой наночастиц щелочных металлов, напыленных на используемую нами подложку. Они могут быть моделированы [2] сплюснутыми по отношению к поверхности эллипсоидами с существенно отличающимися поперечными полуосями. Каждая такая частица имеет два резонанса в спектрах рассеяния и поглощения на разных частотах, соответствующих возбуждению колебаний электронов вдоль каждой из полуосей эллипсоида. Соответствие резонансных пиков поглощения в спектрах экстинкции модам колебаний электронной плазмы вдоль различных полуосей эллиптических частиц и предстояло доказать за счет предложенного принципа определения степени пространственной анизотропии.

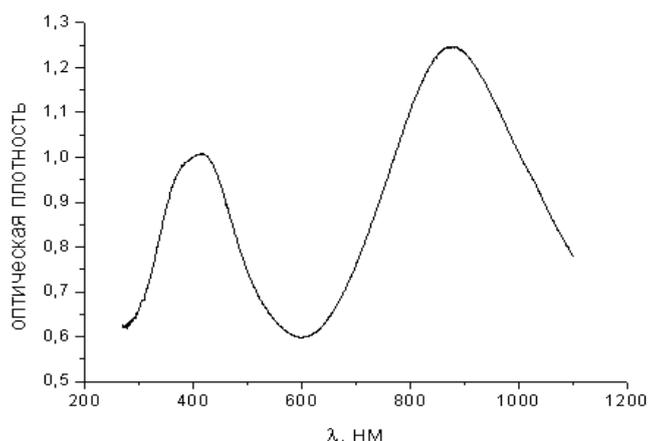


Рис. 2. Спектр экстинкции островковой пленки металлического натрия на сапфире

Спектр экстинкции островковой пленки серебра на поверхности кварца (рис. 3), также исследованной в работе, представлен одиночным пиком поглощения, уширенным вследствие разброса наночастиц по формам. Считается [3], что поглощением света

круглыми частицами образована центральная часть пика, края же резонансного пика возникают благодаря поглощению света более вытянутыми наночастицами. Для проверки данного обстоятельства также была применена поляризационная спектроскопия. Лазер с длиной волны 650 нм попадал в центральную часть резонансного пика, а лазер с длиной волны 532 нм в край полосы поглощения.

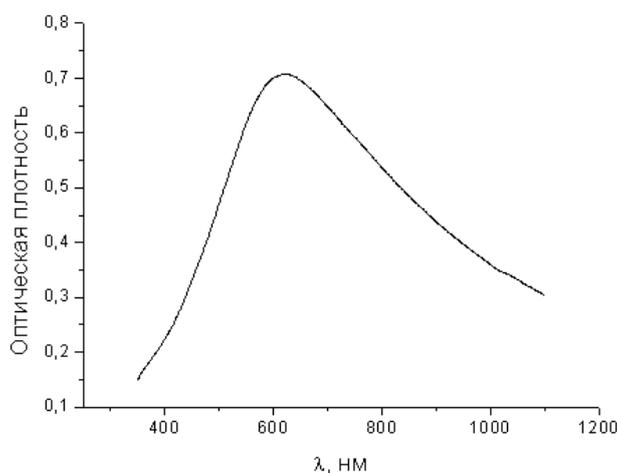


Рис. 3. Спектр экстинкции пленки серебра на кварце

Для сканирования луча по поверхности пленки использовалась механическая система, вращающая кювету с пленкой натрия вокруг своей оси таким образом, чтобы плоскость окна оставалась перпендикулярной направлению распространения лазерного луча, не совпадающему с осью вращения пластинки. В эксперименте скорость вращения пластинки составляла 1,5 оборота в секунду, скорость сканирования луча полупроводникового лазера по поверхности пленки – 2 см/с. Прошедшее через пленку излучение регистрировалось при помощи фотодиода. Полученные сигналы флуктуаций пропускания усиливались и подавались на вход аналогово-цифрового преобразователя с тактовой частотой 25 кГц и в цифровом виде записывались в память компьютера. Каждый цикл измерений продолжался 10 с, что соответствовало набору из 250 000 значений коэффициента поглощения. Статистическая обработка сигнала проводилась на ПК.

Результаты эксперимента и их обсуждение

Целью эксперимента было выяснение возможности выявления степени пространственной анизотропии островковых пленок оптическим методом. Из полученных для каждой пленки 250 000 значений интенсивности проходящего света в результате обработки были определены спектры Фурье полученных сигналов.

Для начала мы проверили предложенный принцип на системе с сильной пространственной анизотропией. Для наночастиц натрия на сапфире (рис. 4) анизотропия оказалась велика настолько, что из данных эксперимента получилось $G \approx 1$ (отношение сигнала на частоте, связанной с вращением луча по поверхности кюветы, к сигналу на частоте $2\omega = 4$ кГц), откуда следует, что $\langle \sigma_1 \sigma_2 \rangle = 0$. Это возможно лишь в том случае, если $\sigma_1 = 0$. Поскольку лазерное излучение с $\lambda = 980$ нм попадало лишь в одну из двух полос поглощения, то это подтверждает гипотезу о соответствии резонансных пиков поглощения модам плазменных колебаний в различных направлениях эллипсоидальных наночастиц.

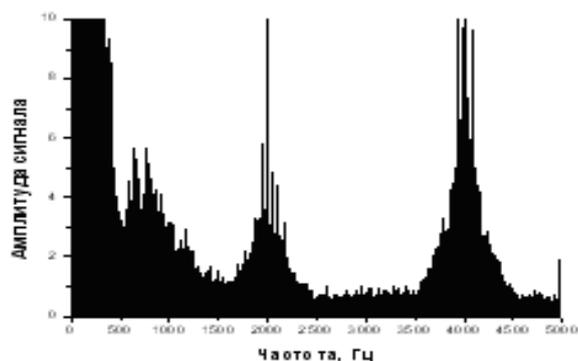


Рис. 4. Фурье-спектр флуктуаций поглощения для наночастиц натрия на поверхности сапфира

Для наночастиц серебра при воздействии на пленку излучения лазера с $\lambda = 650$ нм, совпадающий с центром спектра ее экстинкции мы получили (рис. 5), что анизотропный пик на частоте $2\omega = 2$ кГц много меньше изотропного:

$$I_2V_2 / I_0V_0 = (1 - \langle \sigma_1 \sigma_2 \rangle) / (1 + \langle \sigma_1 \sigma_2 \rangle) \approx 0,$$

поэтому получается $\langle \sigma_1 \sigma_2 \rangle \approx 1$. Это говорит о том, что здесь мы имеем дело с круглыми частицами.

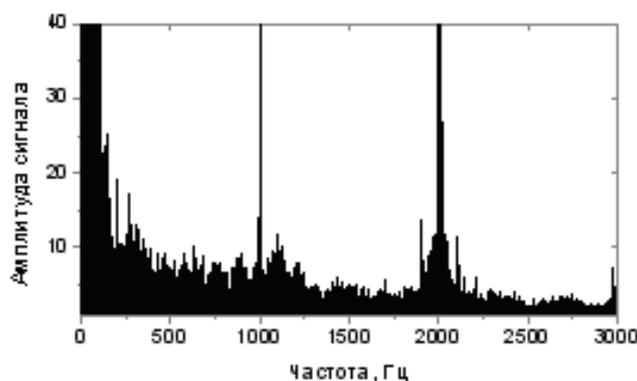


Рис. 5. Фурье-спектр флуктуаций поглощения для наночастиц серебра на поверхности кварца ($\lambda = 650$ нм)

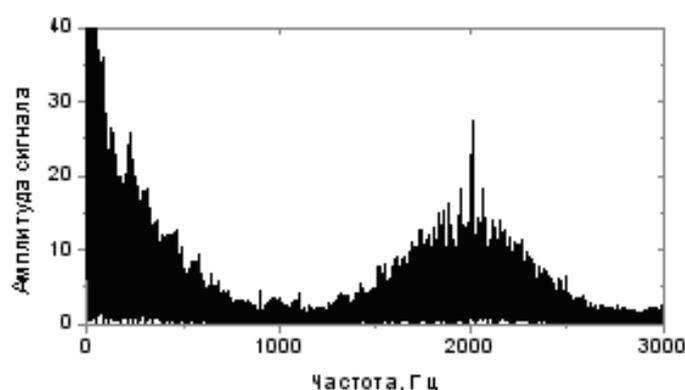


Рис. 6. Фурье-спектр флуктуаций поглощения для наночастиц серебра на поверхности кварца ($\lambda = 532$ нм)

При излучении лазера с длиной волны $\lambda = 532$ нм, которая попадает в коротковолновый край спектра экстинкции пленки, получен существенно отличный Фурье-спектр флуктуаций поглощения (рис. 6). В этой области спектра, как и в пленке натрия, основной вклад в поглощение дают значительно более вытянутые частицы, поскольку здесь

$$I_2V_2 / I_0V_0 = (1 - \langle \sigma_1 \sigma_2 \rangle) / (1 + \langle \sigma_1 \sigma_2 \rangle) \approx 0.5,$$

откуда получается

$$\langle \sigma_1 \sigma_2 \rangle = (1 - I_2V_2 / I_0V_0) / (1 + I_2V_2 / I_0V_0) \approx 1/3.$$

Таким образом, в работе продемонстрирована принципиальная возможность использования довольно простого оптического метода для определения субмикронных параметров металлических островковых пленок и степени анизотропии их оптических свойств, ранее доступных только для электронной и атомно-силовой микроскопии. Можно надеяться, что дальнейшее техническое совершенствование предложенного метода может сделать его надежным инструментом для динамического контроля процессов и структурных изменений, происходящих в островковых пленках при том или ином воздействии на них.

Литература

1. Бонч-Бруевич А.М., Вартамян Т.А., Леонов Н.Б., Пржибельский С.Г., Хромов В.В. Оптический метод измерения структурных параметров островковых пленок // Опт. спектр. – 2000. – Т. 89. – № 438.
2. Вартамян Т.А., Леонов Н.Б., Логунов А.Е., Пржибельский С.Г., Хромов В.В. Связь спектров экстинкции наночастиц металлического натрия с их морфологическими характеристиками. Модификация этих характеристик с помощью процесса фотоатомной эмиссии // Опт. спектр. – 2007. – Т. 102. – №5. – С. 819–824.
3. Гапоненко С.В., Розанов Н.Н., Ивченко Е.Л., Федоров А.В., Бонч-Бруевич А.М., Вартамян Т.А., Пржибельский С.Г. Оптика наноструктур. – СПб: Недра, 2002.

ПЕРЕСТРОЙКА СТРУКТУРЫ НАНОКОМПОНЕНТОВ МОЛЕКУЛЯРНОГО СЛОЯ ПОД ДЕЙСТВИЕМ НАГРЕВА И ОПТИЧЕСКОГО ИЗЛУЧЕНИЯ

А.А. Старовойтов

Научный руководитель – д.ф.-м.н., ст.н.с. Т.К. Разумова

Проведено исследование фото- и термостимулированных процессов в молекулярных слоях полиметиновых красителей на диэлектрических подложках. Дано теоретическое описание структурной и пространственной перестройки наноконструкций слоя.

Введение

Интерес к слоям, содержащим молекулярные компоненты, размеры которых лежат в нанометровой области, связан как с особенностями фотофизических свойств, так и с практическим применением. Молекулярные слои используются в высоких технологиях, в оптоэлектронике и для создания нано- и микроструктур [1]. Необходимо отметить, что молекулярные слои вызывают интерес также благодаря своему особому строению. Для тонких слоев характерна асимметрия взаимодействия отдельных компонентов с окружением (подложкой и соседними молекулами). Это приводит к отличию фотофизических и фотохимических свойств молекулярных слоев от свойств паров, растворов и кристаллов тех же соединений.

Лазерная запись рельефных микроструктур широко используется в современном производстве оригиналов оптических дисков и дифракционных элементов, синтезированных голограмм и т.п. [2]. Были проведены многочисленные эксперименты и теоретические исследования записи и считывания информации с помощью оптических методов. Несмотря на ряд достижений, значительное количество различных проблем, возникающих при записи и воспроизведении информации, остаются нерешенными.

В последние годы растет число исследований, где молекулярные системы применяются для записи информации с помощью фото- [3] или термодеструкции слоя [4]. Фотостимулированный процесс необратимого изменения пространственной ориентации наноконструкций слоев относительно поверхности подложки дает принципиальную возможность записи и считывания долговременной информации оптическими методами без деструкции слоя. Механизм обратимой фотоизомеризации в молекулярных системах открывает возможности для оптической записи на таких средах кратковременной информации. Например, принципы фотореверсивной оптической записи на основе фотостереоизомеризации в пленках изложены в работе [5].

Цель работы – исследование механизмов изменения конформационного состава и пространственной переориентации наноконструкций слоев под действием резонансного фотовозбуждения и нагрева, а также разработка модели перестройки структуры слоя.

Объекты исследования и методики

В качестве объектов исследования были выбраны незамещенные в цепи симметричные дикарбоцианиновые полиметиновые красители. Цепь сопряжения таких молекул связывает две одинаковые гетероциклические концевые группы и содержит нечетное число (5) метиновых групп СН. Молекула является симметричной относительно центрального атома, следовательно, цепь образована прочными выровненными (полуплоскими) углеродными связями [6]. Такой выбор определяется тем, что в гомологическом ряду полиметинов с одинаковыми концевыми группами и разной длиной цепи (n) наибольшие значения сечений оптических переходов наблюдаются у дикарбоцианиновых и трикарбоцианиновых молекул. Большие сечения оптических переходов ($\sigma \sim 10^{-15} \text{ см}^{-2}$)

увеличивают чувствительность спектральных исследований и способствуют эффективному поглощению энергии возбуждающего излучения. Выбор дикарбоцианинов связан с большей фотостойкостью этих соединений по сравнению с более длинноволновыми трикарбоцианинами.

Основные эксперименты проведены на одном из наиболее устойчивых соединений – симметричном полиметиновом красителе кф4011 (1, 3, 3, 1', 3', 3' – гексаметил – 4, 5, 4', 5' – дибензоиндадикарбоцианин – перхлорат). В качестве материала подложки выбрано оптическое стекло – крон К8. Данное решение объясняется тем, что используемый материал прозрачен в спектральных областях поглощения красителя кф4011 и не разрушается под действием лазерного излучения, используемого в экспериментах.

Исследуемые слои наносятся из раствора полиметинового красителя в этиловом спирте на вращающуюся стеклянную подложку. Такая методика обеспечивает круговую симметрию пространственного распределения молекул слоя относительно нормали к поверхности образца. На поверхности стекла существуют отрицательные заряды [7]. В молекулах слоя под действием поверхностных зарядов подложки происходит смещение электронной плотности. Вследствие этого электронное строение химически симметричной молекулы становится асимметричным, и молекулы ориентируются под определенным углом (θ_i) к нормали.

Ориентация полиметиновой цепи мономеров определяет пространственную ориентацию дипольного момента перехода в длинноволновой полосе поглощения (переход $S_0 \rightarrow S_1$). Это дает возможность определять ориентацию молекулы относительно поверхности подложки с помощью исследования анизотропии поглощения слоя в зависимости от угла падения зондирующего линейно-поляризованного излучения. Более подробно данный метод изложен в работе [8].

Спектрально-поляризационные исследования спектров и анизотропии поглощения проводились с шагом 1 нм в диапазоне 300–1100 нм на спектрофотометре СФ-56 с дифракционной решеткой. Диапазон измерения оптической плотности – 0.001–3, точность измерения $\sim 1 \cdot 10^{-4}$. Для поляризационных исследований использовалась призма Глана (скрещенность в диапазоне 400–1000 нм не менее 1/320).

Компонентный состав молекулярного слоя определяется в результате разделения спектра оптической плотности поглощения слоя, полученного при нормальном падении, на спектры отдельных компонентов (F_i) и сопоставления этих спектров с известными спектрами поглощения мономерных изомеров и ассоциатов. При выделении спектров F_i использовались следующие допущения. Длинноволновая полоса поглощения (переход $S_0 \rightarrow S_1$) для каждого из компонентов определяется переходами между нулевым подуровнем основного состояния и тремя электронно-колебательными подуровнями возбужденного состояния (0, 1 и 2), соответствующими одному типу колебаний с частотой $\nu_{кол}$. Таким образом, полоса компонента F_i составлена тремя гауссовыми контурами, соответствующими переходам $0 \rightarrow 0$, $0 \rightarrow 1$ и $0 \rightarrow 2$.

Исследуемые слои содержали следующие конформации – две мономерные формы, различающиеся пространственной конфигурацией: all-trans- и β -cis-стереоизомеры (компоненты F_1 и F_2), а также ассоциированные формы – димеры (F_d) и J-агрегаты (F_j).

Резонансное возбуждение слоя осуществлялось рубиновым лазером (694 нм) и лазером на растворе красителя (727 нм, полуширина ~ 12 нм) в режимах моноимпульса (17 нс) и цуга гигантских импульсов. При термических исследованиях слоев использовалась термостатированная печь, температура в которой контролировалась с точностью $\sim 1^\circ\text{C}$.

В работах [11, 20] указано два типа устойчивой во времени пространственной ориентации компонентов слоев красителя кф4011 (табл. 1). Обнаружена зависимость типа ориентации слоя от поверхностной концентрации молекул. В тонких слоях (количество монослоев $N < 3$) преимущественно существуют компоненты с пространствен-

ной ориентацией первого типа. Второй тип появляется с увеличением толщины слоя и при $N \geq 3$ становится доминирующим. Слои второго типа под действием нагрева [19] и резонансного фотовозбуждения [21] переходят в слой первого типа.

Тип ориентации	$\theta_1 (F_1)$	$\theta_2 (F_2)$	$\theta_d (F_d)$	$\theta_J (F_J)$
	град	Град	град	град
1	68.4	63.8	33.2	75.7
2	56.2	53.5	46.6	63.4

Таблица 1. Типы ориентации и углы наклона наноконфигураций

Результаты исследования

Было установлено, что спектры растворов полиметиновых красителей в смеси вода : этанол (рис. 1, а) существенно ближе к спектру слоев, нанесенных из растворов в этиловом спирте (этанол), см. рис. 1, б. Это объясняется тем, что в смешанном растворе цианинового красителя помимо all-trans-изомера присутствуют и другие компоненты, относительная оптическая плотность поглощения которых зависит от соотношения объемов (V) воды и этанола.

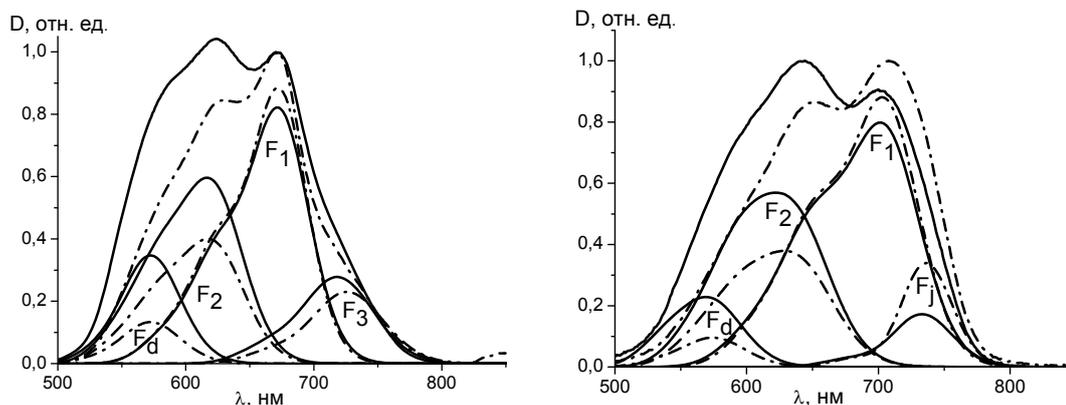


Рис. 1. а) спектры поглощения кф4011 в смеси вода : этанол (сплошная линия – $V(\text{H}_2\text{O}) : V(\text{EtOH}) = 1 : 0,1$; штрихпунктирная – $V(\text{H}_2\text{O}) : V(\text{EtOH}) = 1 : 0,005$); б) спектры поглощения слоя кф4011 и результат их разделения на спектры компонентов (сплошная линия – до облучения; штрихпунктирная – после облучения)

При $V(\text{H}_2\text{O}) : V(\text{EtOH}) = 0.5 : 1$ в спектрах существуют преимущественно полосы поглощения двух мономеров (F_1 и F_2) и длинноволнового компонента (F_3). При увеличении концентрации воды, помимо указанных полос, наблюдается коротковолновое поглощение димера (F_d), за счет чего и происходит дальнейшее уширение спектра поглощения. Увеличение концентрации красителя в смешанном растворе практически не влияет на отношение оптических плотностей мономеров и длинноволнового компонента, но приводит к росту относительной плотности полосы поглощения димера. Это связано с тем, как объяснено в [6], что между молекулярными ионами в димере создаются водородные связи, если не непосредственно, то с помощью включения в димер молекул H_2O .

Как резонансное фотовозбуждение, так и нагрев вызывают преобразование строения молекулярного слоя, которое обусловлено двумя процессами. Происходит взаимная перестройка и мономерных, и ассоциированных форм, изменяющая компонентный состав слоя (рис. 1, б). Механизм структурной перестройки связан с распадом димеров на мономеры и объединением димеров в J-агрегаты, а также со стереоизомер-

ризацией мономеров. Одновременно с этим наноконпоненты слоя переориентируются в пространстве.

Первый процесс наблюдается при сравнительно малых величинах плотности энергии светового импульса E_s и при температурах нагрева более 45 °С. Для перестройки пространственной ориентации требуются существенно большие величины как E_s (табл. 2), так и температуры нагрева (табл. 3).

№	λ_{ex} , нм	режим	E_s , мДж/см ²	E_{Sab} , мДж/см ²	$E_{Sab}/N_s \times 10^{-14}$, мДж/мон.	$\Delta\theta_1$, град	$\Delta\theta_2$, град	$\Delta\theta_d$, град	$\Delta\theta_J$, град
1	694.3	моноимп.	19	0.97	0.23	3.9	0	0	5.1
2	694.3	моноимп.	30	1.28	0.33	7.5	2.9	-5.1	10.4
3	694.3	моноимп.	42	1.98	0.53	8.8	8.8	-7.1	11.7
4	694.3	пачка, 6 имп.	13	1.03	0.16	10.9	6.5	-10	9.5
5	727	моноимп.	30	2.16	0.28	6.7	0	0	7
6	727	моноимп.	100	7.76	0.87	7.2	2.9	-5.5	7
7	727	пачка, (2-3) имп.	20	0.72	0.33	8.9	3.3	-2.7	7.2

Таблица 2. Влияние условий возбуждения на перестройку углов ориентации компонентов слоя

T ⁰ С	t, мин	NF ₁ /N _s	NF ₂ /N _s	NF _d /N _s	NF _J /N _s	Ns0 * 10 ⁻¹⁴	θ_1 , град	θ_2 , град	θ_d , град	θ_J , град
20	0	0.17	0.23	0.28	0.32	4.19	61.8	56.4	42	67.6
150	60	0.2	0.14	0.27	0.39		62.6	57.8	44.8	70.5
20	0	0.18	0.20	0.31	0.31	4.17	60.3	55.5	43.2	68.
175	25	0.21	0.18	0.20	0.41		65.5	60	40.8	71.7
20	0	0.19	0.21	0.35	0.25	4.8	60.7	54.6	45.2	70.2
225	10	0.16	0.12	0.24	0.49		68.1	58.2	39.9	74.4

Таблица 3. Перестройка молекулярных форм и углов ориентации компонентов слоя при нагреве

Как взаимное преобразование наноконпонентов, так и изменение их пространственной ориентации приводят к изменениям спектра поглощения и поляризационных характеристик молекулярного слоя.

Рассмотрим факторы, которые определяют предельные углы и характер пространственной перестройки. Изучено влияние на процесс переориентации режима возбуждения, исследованы моноимпульсный режим, цуг гигантских импульсов и свободная генерация. На рис. 2, а, представлена зависимость углов ориентации от числа моноимпульсов лазера на красителе ($\lambda = 727$ нм) и суммарной энергии возбуждения при плотности энергии возбуждения 100 мДж/см².

Величины углов мономеров и J-агрегата растут, а угол наклона димеров уменьшается. Эти зависимости имеют насыщающийся характер. Обнаружено, что при малых плотностях энергии возбуждения изменения углов ориентации наблюдаются только для возбуждаемого светом all-trans-изомера и J-агрегата, при этом предельно достижимые величины изменения углов меньше. На рис. 2, б, представлена зависимость предельно достижимых углов ориентации мономеров от поверхностной плотности энергии

моноимпульса рубинового лазера. Чем больше плотность энергии в моноимпульсе, тем больше предельно достижимые изменения угла ориентации (см. рис. 3), и тем быстрее выход на предельный угол насыщения. Эта зависимость также имеет насыщающийся характер.

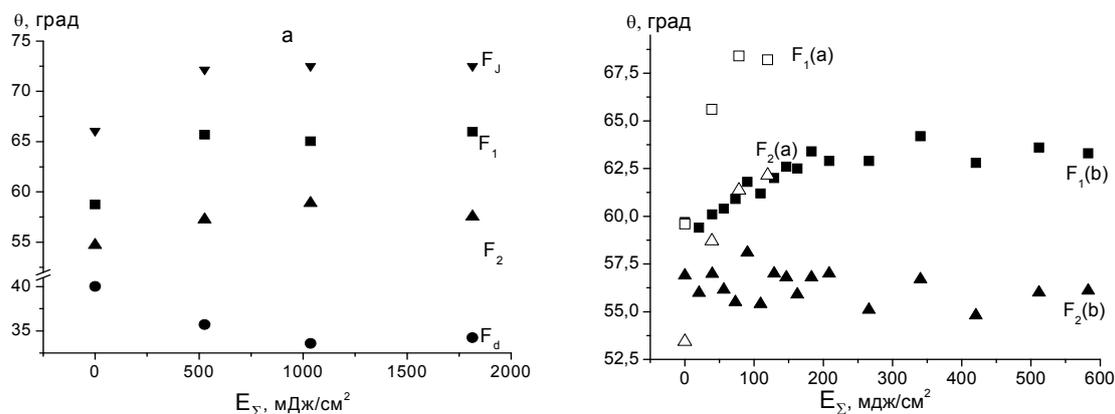


Рис. 2. Углы θ_i компонентов в зависимости от суммарной плотности энергии E_s : а) $\lambda=727$ нм, $E_s = 100$ мДж/см²; б) $\lambda=694$ нм, $E_s = 42$ (а) и 19 (б) мДж/см²

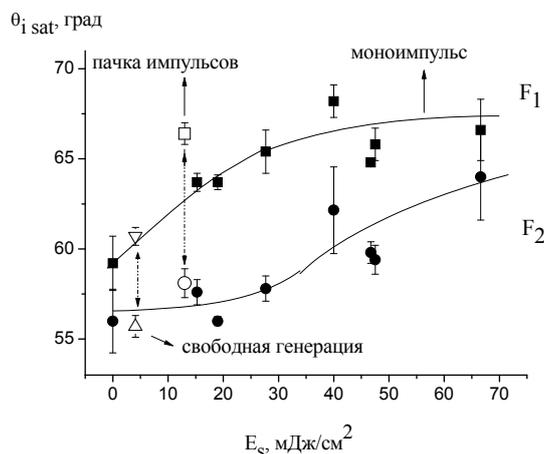


Рис. 3. Предельные углы ориентации $\theta_{i \text{ sat}}$ в зависимости от плотности энергии в импульсе при разных режимах генерации рубинового лазера

Как видно из рис. 3, пачка из 6 импульсов действует более эффективно. Можно предположить, что наблюдаемое отличие в закономерностях изменения угла при разных режимах фотовозбуждения связано с разной плотностью как энергии, так и мощности возбуждения, определяемой разной длительностью отдельных импульсов, генерируемых лазером в разных режимах, и с разными временными интервалами между ними.

Результаты воздействия зависят также от длины волны (см. рис. 2), что связано, по-видимому, с разным количеством поглощенной энергии и с изменением относительной роли фото- и термостимулированных процессов в основном и возбужденном состояниях компонентов. Установлено, что при моноимпульсном возбуждении с плотностью энергии 30 мДж/см² при длине волны возбуждения 727 нм переориентация затрагивает только all-trans-изомер и J-агрегат, в то время как при возбуждении с 694 нм переориентируются все компоненты.

Из табл. 2 видно, что эффективность возбуждения рубиновым лазером существенно выше: при меньших величинах поглощенной энергии величины изменения углов ориентации заметно больше, чем в случае возбуждения с 727 нм.

Модель перестройки структуры

Модель пространственной перестройки компонентов слоя должна удовлетворительно описывать экспериментальные закономерности как термо-, так и фотопереориентации. Процесс термической перестройки слоев полиметинового красителя кф4011 в диапазоне температур (120–230)°С хорошо описывается моделью, включающей как термостимулированные изменения ориентации молекул, находящихся в нижнем состоянии S_0 , так и перестройку их окружения [9]. Удовлетворительное совпадение экспериментальных и расчетных данных получается для модели, представленной на рис. 4 (нижняя кривая, основное состояние S_0).

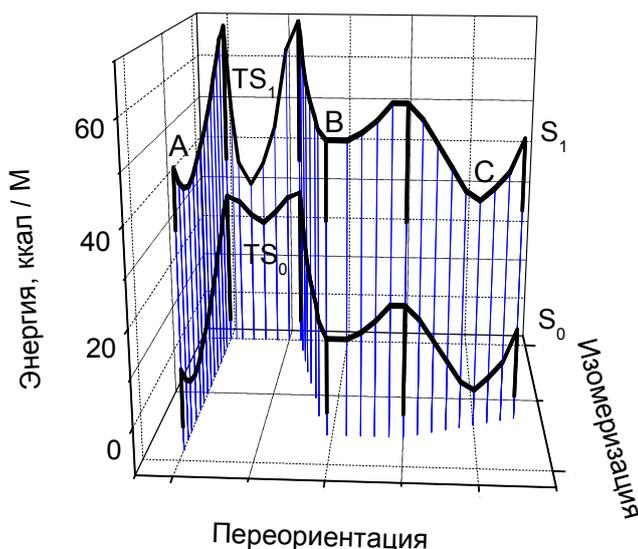


Рис. 4. Энергетическая модель термо- и фотоиндуцированных перестроек пространственной ориентации мономера F1

Модель включает молекулы трех типов пространственной ориентации – исходная молекула A с ориентацией второго типа (см. табл. 1) и два типа молекул с перестроенной ориентацией, отличающихся только степенью перестройки окружения, – промежуточный тип (B) и молекулы с первым типом ориентации (окончательный тип C). В этой модели первой ступенью перестройки ориентации является термостимулированный разворот фрагмента молекулы мономера на угол $\sim 90^\circ$ вокруг С-С связи полиметиновой цепи (начальный этап термоизомеризации, развернутое состояние TS_0).

Этот поворот сопровождается незначительным изменением пространственной ориентации молекулы и перестройкой ее окружения, что приводит к некоторому падению энергии и образованию минимума TS_0 на барьере изомеризации в состоянии S_0 (нижняя кривая на рис. 4). Из этого минимума изогнутая молекула может либо вернуться в исходную ориентацию A, либо закончить пространственную переориентацию и перейти в промежуточно перестроенную ориентацию B с последующим переходом в конечный тип перестройки – C.

Модель фотоперестройки, возникающей вследствие резонансного возбуждения мономера F_1 слоя и перехода исходной молекулы A в возбужденное состояние S_1 , представлена на рис. 4 (верхняя кривая). При построении модели фотоизомеризации слоя использованы результаты исследования фотоизомеризации ПК в растворах [10]. Близость положений максимумов спектров поглощения all-trans-изомера в слоях и растворах ПК показывает, что относительные положения энергетических термов молекул мало изменяются при изменении окружения. Отличия связаны лишь с большим неоднородным уширением спектра поглощения в слое.

Изменение окружения преимущественно влияет на величину энергии активации изомеризации (E_a), определяемую суммой энергии активации внутримолекулярных процессов поворота фрагмента и энергии активации преодоления сопротивления окружения молекулы. Вследствие увеличения жесткости среды энергия активации в основном состоянии F_1 в слое (E_{a0}) увеличивается почти в два раза по сравнению с E_{a0} all-trans-изомера в растворе. При оценке параметров фотоизомеризации слоя предполагалось, что величины доли энергии активации, связанной с преодолением реакции окружения, одинаковы в основном и возбужденном состоянии F_1 . В этом случае разность величин активации изомеризации в основном и возбужденном (E_{a1}) состоянии слоя определяется только разностью энергий активации внутримолекулярных процессов поворота. Близость спектральных параметров all-trans-изомера в растворе и слое позволяет предположить, что величины ($E_{a0} - E_{a1}$) для состояний S_1 и S_0 в слое и растворе близки. В рамках этих предположений, при использовании параметров термоизомеризации для нижнего состояния слоя и данных для раствора была построена энергетическая модель фотостимулированной перестройки слоя полиметинового красителя, приведенная на рис. 4 (верхняя кривая).

Переход молекулы мономера в S_1 уменьшает энергию активации разворота фрагмента молекулы вокруг связи C-C цепи (фотоизомеризация). Теоретические расчеты свободной энергии состояния S_1 молекулы [10] показали, что поворот фрагмента молекулы сопровождается изменением внутримолекулярного распределения заряда и свободной энергии молекулы. Поворот фрагмента приводит вначале к росту энергии возбужденной молекулы и к образованию барьера изомеризации. Затем, по мере приближения к углу 90° , энергия развернутой молекулы падает (глубокий минимум TS_1 на барьере изомеризации в развернутом состоянии). Следствием такого падения энергии является уменьшение величины барьера, а также рост вероятности безызлучательного перехода $TS_1 \rightarrow TS_0$.

Таким образом, как при термо-, так и при фотостимулированной переориентации слоев первым этапом перестройки является процесс изомеризации в основном или в возбужденном состояниях, переводящий молекулу F_1 в долгоживущий минимум TS_0 . Время жизни этого минимума зависит от температуры. При комнатной температуре выход перехода из состояния TS_0 в основное состояние \underline{B} близок к 95%.

Существенным отличием процесса фотоперестройки является меньший барьер изомеризации в S_1 состоянии, а также увеличение запаса колебательной энергии молекул слоя вследствие трансформации части поглощенной световой энергии в тепло, что также приводит к увеличению вероятности изомеризации и перехода в TS_0 . Скорость прироста температуры зависит от населенности основных и возбужденных состояний, от числа молекул, перешедших в долгоживущую яму TS_0 , и от скорости диссипации тепловой энергии слоя. Вследствие этого скоростной режим нагрева слоя определяется плотностью мощности и энергии излучения, длительностью одиночных импульсов, их числом в цуге и длительностью цуга. Таким образом, модель, изображенная на рис. 4, хорошо описывает результаты экспериментальных исследований термо- и фотопереориентации молекулярных слоев.

Заключение

Проведено исследование фото- и термостимулированных процессов в молекулярных слоях полиметиновых красителей на диэлектрических подложках. Полученные в работе результаты позволяют сделать ряд выводов.

(1) Фотостимулированные процессы перестройки структуры и пространственной ориентации наноконструкций молекулярного слоя приводят к существенному изменению оптической плотности и спектров поглощения слоя. Наиболее сильные изменения наблюдаются для случая косого падения линейно поляризованного зондирующего излучения.

(2) Величины предельного изменения углов ориентации зависят от спектрального диапазона, режима возбуждения и плотности энергии излучения в импульсе (E_S). Наиболее эффективно резонансное возбуждение мономерной all-trans-формы. Максимальное изменение углов ориентации составляет $\sim 15^\circ$. При возбуждении моноимпульсным излучением рубинового лазера для достижения максимальных углов перестройки требуется $E_S \geq 45$ мДж/см². При режиме возбуждения пачкой гигантских импульсов необходимая величина E_S понижается.

(3) Предложена энергетическая модель, которая описывает термо- и фотостимулированные процессы в основном и возбужденном состояниях. Определяющий механизм изменения углов ориентации молекулярных компонентов слоев полиметиновых красителей – стереоизомеризация компонентов слоя в возбужденном и основном состояниях.

(4) Процессы пространственной перестройки компонентов молекулярных слоев могут быть использованы для оптической записи и считывания информации. Наиболее эффективная спектральная область записи расположена вблизи максимума поглощения all-trans-изомера. Наибольшая чувствительность считывания соответствует области перекрытия спектров димера и β -cis-изомера.

Для окончательных выводов о перспективности практического использования исследованных объектов необходимо проведение дальнейших исследований.

Литература

1. Nanomarkets Co. (2007). Nanostorage: The impact of nanotechnology on memory, disk, drives and other storage devices. A technology and market assessment for 2004-2011. USA, Glen Allen. – Режим доступа: <http://www.nanomarkets.net>
2. Оптические диски: история, состояние, перспективы развития / В.В. Петров, А.А. Крючин и др.; под ред. А.Г. Додонова. – Киев: Наукова думка, 2004. – 26 с.
3. Yang Wang, Donghong Gu, Fuxi Gan. Application of subphthalocyanine thin films in high-density optical recording // Sixth International Symposium on Optical Storage. Proceedings of SPIE. – 2003. – V. – 5060. – P. 15–19.
4. Kravets V.G., Vinnichenko K.L., Prygun O.V. Characterization and optical properties of organic dye films as recording media // Semiconductor Physics, Quantum Electronics & Optoelectronics. – 2000. – V.3 (№4). – P. 520–522.
5. Фотореверсивная оптическая запись в пленках аморфных азосодержащих полимеров / А.Н. Симонов, Д.В. Ураев, В.П. Шибяев, С.Г. Костромин. Квантовая электроника. – 2002. – Т. 32 (№2). – С. 143–148.
6. Теренин А.Н. Фотоника молекул красителей и родственных органических соединений. – Л: Наука, 1967. – 616 с.
7. Quantification of methylene blue aggregation on a fused silica surface and resolution of individual absorbance spectra / S. M. Ohline, S. Lee, S. Williams, C. Chang // Chemical Physics Letters. – 2001. – V. 346. – P. 9–15.
8. Калитиевская Е.Н., Крутякова В.П., Разумова Т.К. Исследования влияния толщины слоя дикарбоцианиновых красителей на конформационный состав и пространственную ориентацию компонентов слоя // Оптика и спектроскопия. – 2004. – Т. 97 (№6). – С. 960–968.
9. Калитеевская Е.Н., Крутякова В.П., Разумова Т.К. Изменение конформационного состава и пространственной ориентации молекулярных компонентов слоя дикарбоцианинового красителя в результате нагрева // Оптика и спектроскопия. – 2005. – Т. 99 (№6). – С. 1012–1019.
10. Батюто Я.В., Разумова Т.К., Тарновский А.Н. Ступенчатая фотоизомеризация дикарбоцианиновых красителей // Оптика и спектроскопия. – 2002. – Т. 93(№3). – С. 434–442.

**Научная школа
«ТЕХНОЛОГИИ
ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ
ВЫЧИСЛЕНИЙ И СИСТЕМ»**

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГЕОМЕХАНИЧЕСКИХ ПРОЦЕССОВ

А.И. Уланов (Санкт-Петербургский государственный политехнический университет)
Научный руководитель – д.т.н., профессор Ю.Я. Болдырев
(Санкт-Петербургский государственный политехнический университет)

В работе исследуется возмущение напряженного состояния двухфазной среды в окрестности нефтяной скважины при бурении, оценивается возможность относительного скольжения вдоль границ различных слоев породы. Цель работы – разработка математического аппарата и создание вычислительного пакета для моделирования геомеханических процессов.

1. Введение

Геомеханика – наука, изучающая физико-механические свойства пород, их напряженное состояние, процессы деформирования и разрушения, происходящие под влиянием природных и технологических факторов. Теория насыщенных пористых сред является одним из важнейших разделов геомеханики. В данной работе рассматривается математическая модель, позволяющая исследовать поведение насыщенного жидкостью или газом грунта в различных геотехнических процессах, таких как:

- проседание поверхности в результате разработки нефтяных или газовых резервуаров (рис. 1);
- исследование напряженного состояния грунта при бурении скважины;
- исследование процессов консолидации грунта;
- моделирование сдвига слоев грунта друг относительно друга.

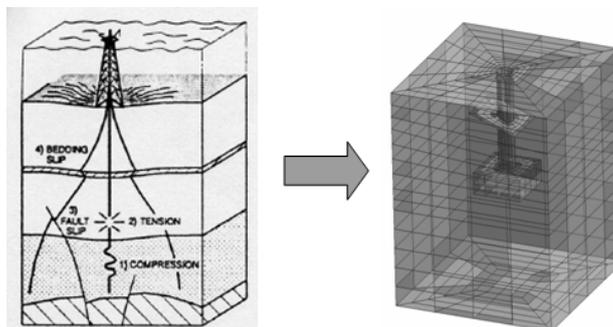


Рис. 1. Пример модели резервуара

В работе используется модель, описывающая совместный процесс определения напряженного состояния и диффузии порового давления жидкости в пороупругой насыщенной среде [1]. Исследования проводятся в предположении, что рассматриваемое тело состоит из двух частей, разделенных зоной контакта. Отличительной особенностью рассмотренной проблемы является использование контактного слоя, с помощью которого проводилось моделирование эффекта скольжения слоев грунта. Такой подход к описанию и реализации процессов, протекающих в рассматриваемом теле, вносит существенные изменения и усложнения в модель двухфазной пороупругой среды. Для анализа возникновения и развития скольжения был использован закон Мора-Кулона. Проведены расчеты, определяющие влияние тех или иных свойств рассматриваемого тела на поведение процесса скольжения. Для численной реализации данной задачи использовался метод конечных элементов.

Следует отметить, что задачи подобного плана не входят в состав ни одного из существующих на сегодняшний день вычислительных комплексов (ANSYS, FLAC 3D, ECLIPSE, PATRAN и др.).

2. Постановка задачи

2.1. Основные уравнения и тождества

Пористая среда моделировалась как совокупность движущейся жидкости, окружающего ее порового пространства и деформируемого остова. Для описания поведения этой среды были приняты следующие предположения [5]:

- протекание жидкости через пористую среду осуществляется в соответствии с законом Darcy;
- жидкость сжимаема;
- пористая среда изотропна;
- перемещения остова малы;
- пористая среда линейна.

Запишем уравнения, описывающие процесс деформации среды и движения жидкости в порах исследуемого тела, в том виде, в каком они были получены (Biot, 1941; Boutéca, 1992; Coussy, 1995; Lewis and Schrefler, 1998). Уравнения диффузии порового давления жидкости и для перемещений будут иметь вид [1]:

$$M \frac{\partial p}{\partial t} - \frac{k}{\mu} \Delta p = -b \frac{\partial}{\partial t} \operatorname{div} \mathbf{u} . \quad (1)$$

$$G \Delta \mathbf{u} + \gamma \nabla \operatorname{div} \mathbf{u} = b \nabla p . \quad (2)$$

В уравнениях (1), (2) использованы обозначения: p – поровое давление жидкости; \mathbf{u} – вектор перемещения точек твердого тела; k – коэффициент проницаемости; μ – вязкость насыщающей жидкости; G и γ – упругие постоянные, характеризующие деформируемый скелет; b – константа Био.

2.2 Вариационная формулировка

Рассмотрим область Ω в пространстве R^m ($m = 2$ или 3), с границей S и внешней нормалью \mathbf{n} . Пусть \mathbf{u} и \mathbf{q} – функции из пространства $(W_2^1(\Omega))^3$. Для заданных таким образом функций постулируем правильность следующих выражений:

$$\int_{\Omega} \Delta \mathbf{u} \cdot \mathbf{q} d\Omega = - \int_{\Omega} (\nabla \mathbf{u})^T : \nabla \mathbf{q} d\Omega + \int_S \frac{\partial \mathbf{u}}{\partial n} \cdot \mathbf{q} ds . \quad (3)$$

$$\int_{\Omega} \nabla f \cdot \mathbf{q} d\Omega = - \int_{\Omega} f \operatorname{div} \mathbf{q} d\Omega + \int_S f \mathbf{q} \cdot \mathbf{n} ds . \quad (4)$$

где $\frac{\partial \mathbf{u}}{\partial n} = \frac{\partial \mathbf{u}}{\partial \mathbf{r}} \mathbf{n} = (\nabla \mathbf{u})^T \mathbf{n}$. Умножим выражение (2) на вектор-функцию \mathbf{q} , принадлежащую тому же классу, что и \mathbf{u} . Используя формулы (3)–(4), полученное выражение проинтегрируем по области Ω и сгруппируем слагаемые, относящиеся к вычислению интегралов по Ω и по S :

$$\int_{\Omega} (G(\nabla \mathbf{u})^T : \nabla \mathbf{q} + \gamma \operatorname{div} \mathbf{u} \operatorname{div} \mathbf{q} + b \nabla p \mathbf{q}) d\Omega - \int_S (G(\nabla \mathbf{u})^T + \gamma \operatorname{div} \mathbf{u} \mathbf{I}) \mathbf{n} \cdot \mathbf{q} ds = 0 . \quad (5)$$

Слагаемые под контурным интегралом представляют собой нормальное и касательные напряжения на границе области S , которые при данной вариационной формулировке уравнения (2) являются естественными граничными условиями.

Применив аналогичные преобразования для уравнения диффузии порового давления (1), получим:

$$\int_{\Omega} \left(M \frac{\partial p}{\partial t} q + \frac{k}{\mu} \nabla p \cdot \nabla q + b \frac{\partial \operatorname{div} \mathbf{u}}{\partial t} \cdot q \right) d\Omega - \oint_S \frac{k}{\mu} \nabla p \mathbf{n} q ds = 0. \quad (6)$$

Как и в предыдущем случае, выражение под контурным интегралом представляет естественное краевое условие для давления.

Интегральные тождества (5) и (6) являются основой для получения конечно-элементной модели рассматриваемой задачи.

2.3. Модель контактного слоя

Для моделирования контактного слоя воспользуемся идеей интерфейсного элемента, предложенного в работе [9]. Рассмотрим две области Ω_1 и Ω_2 с границами S_1 , S_2 соответственно (рис. 2). Через S_c обозначим участок общей границы (поверхность контакта). Оставшуюся часть границы обозначим через $S = S_1 \cup S_2 \setminus S_c$.

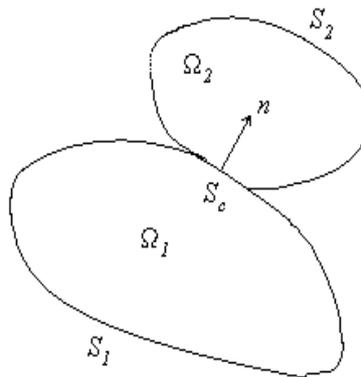


Рис. 2. Иллюстрация контакта

При составлении общей системы уравнений метода конечных элементов, соответствующей уравнениям (1) и (2) в областях Ω_1 и Ω_2 , были использованы интерфейсные элементы, расположенные вдоль контактного слоя S_c .

Предполагалось, что контактный слой имеет бесконечно малую толщину и представляет собой плоскую поверхность. К характеристикам слоя отнесем проницаемость D и жесткость C . Жесткость – отношение напряжения к соответствующей деформации: чем больше требуется усилий (напряжения) для получения заданной деформации, тем больше жесткость (способность элемента конструкции противостоять упругому отклонению). В рассматриваемой задаче матрица жесткости C задавалась как [9]

$$C = \begin{pmatrix} C_s & 0 & 0 \\ 0 & C_s & 0 \\ 0 & 0 & C_n \end{pmatrix}. \quad (7)$$

Набор характеристик интерфейсного элемента [9] был дополнен параметром D , который при достаточно больших значениях обеспечивает непрерывность порового давления при переходе через контактный слой.

2.4. Моделирование скольжения

Среди процессов, возникающих на контактном слое, следует выделить возможность возникновения при определенном напряженном состоянии открытия и проскальзывания в зоне контакта. Рассмотрим последовательность расчета напряженного состояния тела при моделировании процесса скольжения одного берега контакта относи-

тельно другого. Предположим, что нижняя часть исследуемого объекта (рис. 3) закреплена: перемещения $u = v = w = 0$. Под действием приложенной нагрузки \mathbf{F} (рис. 3) в контактном слое возникают касательные напряжения, и если нарушается закон Кулона-Мора (8), возникает скольжение.

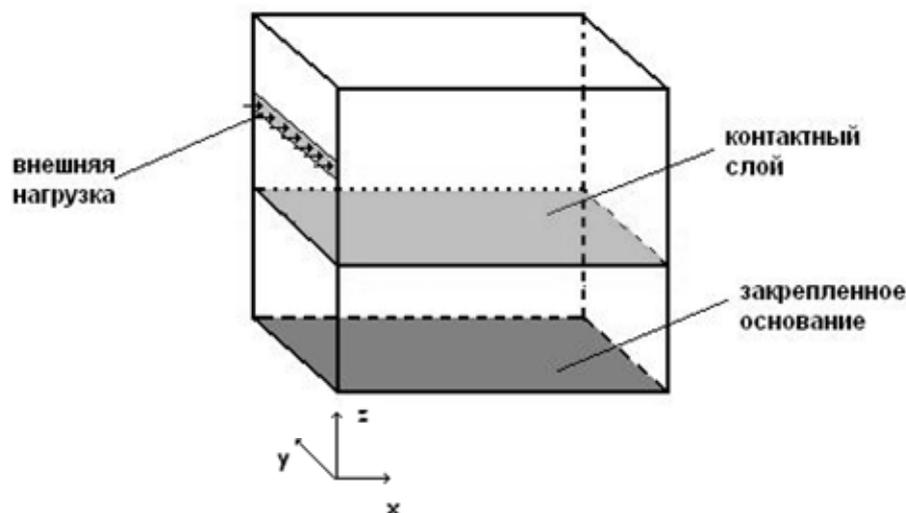


Рис. 3. Исследуемое тело

В качестве критерия скольжения используем закон Mohr-Coulomb:

$$|\sigma_s| \leq K \cdot \sigma_n + C_H. \quad (8)$$

где введены обозначения: σ_n – нормальное напряжение на поверхности контактного элемента; σ_s – касательное напряжение; K – коэффициент трения; C_H – выбранное постоянное напряжение сцепления (в дальнейшей терминологии обозначается через Coh – когезия).

В случае возникновения условий скольжения меняются свойства контактного элемента: полагаем равенство нулю компонент матрицы жесткости, отвечающих за касательные перемещения ($C_s=0$), тем самым, позволяя независимое перемещение точек верхнего и нижнего берегов слоя. В соответствии с приведенным описанием, для моделирования скольжения использовался следующий алгоритм:

- 1) производился расчет деформированного состояния пороупругой среды;
- 2) в соответствии с законом (8) находятся элементы контактного слоя, где выполнялось условие скольжения;
- 3) производится расчет деформированного состояния с учетом поправок, возникающих в матрице элемента контактного слоя, на котором выполнилось скольжение.

Шаги 2, 3 повторялись до тех пор, пока на очередной итерации в контактном слое не появятся новые «скользящие» элементы, или пока весь контактный слой не уйдет в скольжение.

3. Результаты

В данной работе расчеты напряженного состояния образца и процесса диффузии порового давления в нем проводились при различных условиях. Одним из основополагающих критериев для разделения результатов на различные группы является наличие или отсутствие скольжения. В первую группу были выделены расчеты, при которых моделировались варианты решения без учета скольжения. В свою очередь, эти результаты можно разделить на два основных класса в зависимости от условий, задаваемых для давления:

- давление фиксировано (линейное распределение или постоянное во всем объеме), рассчитывается напряженное состояние твердого тела, соответствующее механической нагрузке и заданному полю давления,
- нестационарная задача расчета напряженного состояния и диффузии порового давления до установления.

Сравнивались решения, полученные при фиксированном распределении давления в области и в результате завершения процесса установления.

Вторая группа расчетов описывает эффект скольжения. Эксперименты данного раздела разбиты на части, моделирующие поведение исследуемого объекта в зависимости от различных факторов и комбинаций условий. Кратко перечислим основные принципы и критерии, по которым производилось разделение результатов:

- влияние величины когезии (Coh) на количество элементов, для которых выполняется критерий скольжения;
- влияние проницаемости;
- влияние величины давления в задаче установления;
- влияние силы трения.

Дальнейшие результаты приводятся в безразмерных величинах.

Детально остановимся на расчетах, описывающих влияние проницаемости. Для этого будем решать задачу в следующей постановке:

- в качестве исследуемого тела будем рассматривать куб с горизонтально расположенным контактным слоем (рис. 3);
- решается нестационарная задача установления постоянного распределения давления;
- выбирается фиксированное значение когезии Coh ;
- варьируется проницаемость k (1).

Для этого на верхней грани куба задавалось постоянное давление. Производился нестационарный расчет при нулевом начальном приближении (рис. 4). Результаты этой серии экспериментов приведены в таблице. Все элементы, ушедшие в скольжение, были получены на первом временном шаге.

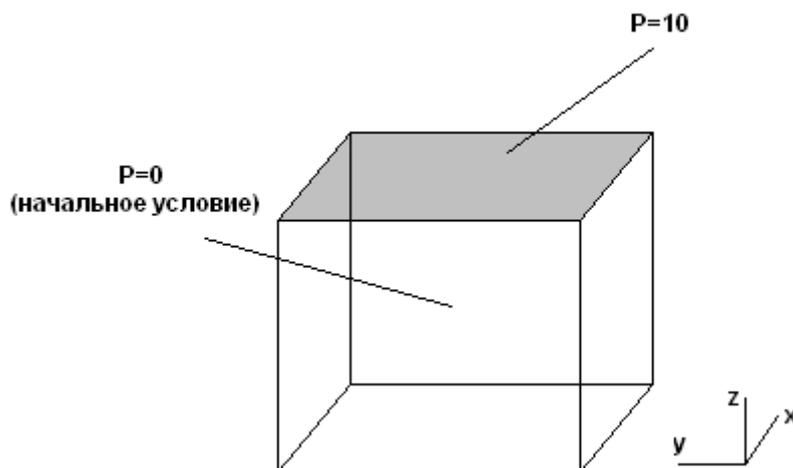


Рис. 4. Условия для давления

Скольжение элементов контактного слоя обусловлено наличием внешней нагрузки, действующей в направлении оси X (рис. 3). Элементы, на которых возникло скольжение, отмечены на рис. 5.

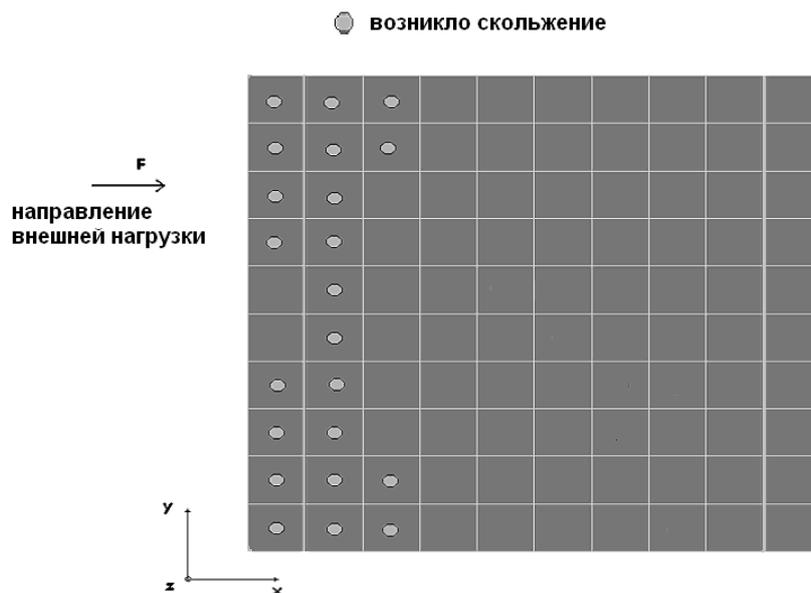


Рис. 5. Контактный слой, вид сверху

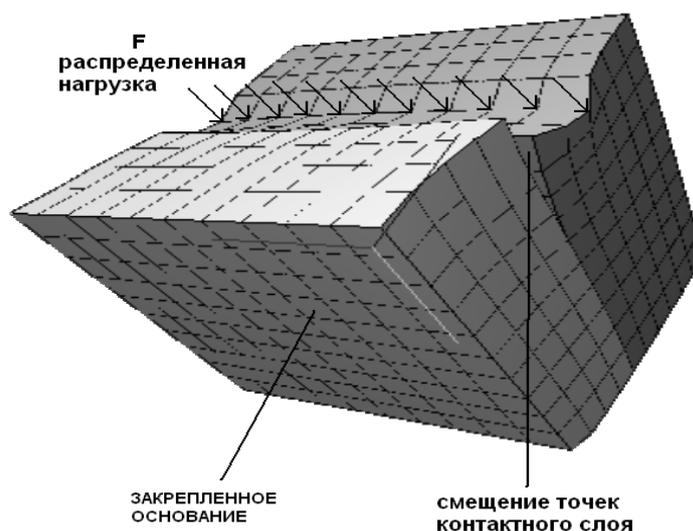


Рис. 6. Общий вид деформации исследуемого тела

Градиент порового давления в процессе установления будет тем больше, чем меньше значение коэффициента проницаемости k . Соответственно изменяется (возрастает при уменьшении k) правая часть уравнения (2), что эквивалентно появлению новых нагрузок. Таким образом, изменяется общее напряженно-деформированное состояние тела, что приводит к появлению большого количества элементов, на которых выполняется критерий скольжения.

Значение k	Coh	Количество элементов
0.1	2.7	22
0.2	2.7	8
0.3	2.7	4
1	2.7	4

Таблица. Исследование влияние k в задаче установления постоянного давления

4. Заключение

Выполненные в работе расчеты моделируют поведение пороупругой насыщенной среды и возможного скольжения на контактном слое в соответствии с законом Кулона-Мора с учетом воздействия распределенной внешней нагрузки. Для описания поведения жидкости в исследуемом теле необходимо вводить дополнительное уравнение, что существенно усложняет исследуемую задачу: наряду с уравнениями деформации, используемыми в теории твердого тела, необходимо рассматривать уравнение диффузии порового давления жидкости.

Проведено исследование влияния различных параметров на возникновение и развитие процесса скольжения:

- влияние величины когезии (Coh) на количество элементов, для которых выполняется критерий скольжения;
- влияние значения проницаемости на возникновение процесса скольжения;
- влияние величины давления в нестационарной задаче установления;

Перспективным направлением при усовершенствовании алгоритма расчета является создание многопроцессорной версии данной программы. Такой подход, наряду с использованием технологии хранения редких матриц (SPARSE), может дать значительный результат в дальнейших исследованиях.

Литература

1. O.C. Zienkiewicz, A.H.C. Chan, M. Pastor, B.A. Schrefler, T. Shiomi. // Computational Geomechanics, 1999.
2. Костерин А.В. Новые модели и обобщенное решение нелинейных задач механики насыщенных пористых сред. // Математическое моделирование. – 2001. – Т.13. – №2.
3. Николаевский В.Н. Механика пористых трещиноватых сред. – М. Недра, 1984. – 232 с.
4. Alexander H.-D. Cheng Fundamental of Poroelasticity. // Analysis and Design Method. – 1993. – Vol 2. – P. 113–171.
5. M. Mainguy and P. Longuemare. Coupling Fluid and Rock Mechanics. Oil and Gas Science and Technology. // Rev. IFP. – 2002. – Vol. 57. – P. 355–367.
6. G. Gambolati, M. Ferronato. Finite element analysis of land subsidence. // Int. J. Anal. Meth. Geomech. – 2001. – 25. – P. 307–327.
7. Фадеев А.Б. Метод конечных элементов в геомеханике. – М. Недра, 1987. – 221 с.
8. Зенкевич О., Чанг И. Метод конечных элементов в теории сооружений и в механике сплошных сред. – М. Недра, 1974. – 240 с.
9. R.E. Goodman, R.L. Taylor, T.L. Brekke A Model for the Mechanics of Jointed Rock. // J. Soil Mech. and Found. Division. –1968, May. – P. 637–659.

РАСЧЕТ АКУСТИЧЕСКИХ ХАРАКТЕРИСТИК АВТОМОБИЛЬНЫХ ЗЕРКАЛ

К.Ю. Замогин

(Санкт-Петербургский государственный политехнический университет)

Научный руководитель – к.т.н., доцент Ю.К. Шиндер

(Санкт-Петербургский государственный политехнический университет)

Целью работы являлась разработка инструмента, позволяющего получать акустические характеристики объекта, обдуваемого воздушным потоком, средств, которые позволили бы рассчитать аэроакустические свойства автомобильных зеркал (и не только) различной формы, провести сравнение расчетных данных с экспериментальными.

1. Введение

Автопроизводителям давно известно, что нерегулярная область потока за внешними боковыми зеркалами дальнего вида автомобиля является существенным источником шума как внутри, так и снаружи транспортного средства во время его движения. Собственно форма корпуса зеркального элемента, его положение в потоке, угол к поверхности боковой двери автомобиля – все эти факторы влияют на аэроакустические свойства автомобиля.

Методы решения задач акустики, возникающих при обтекании движущихся объектов (самолетов, автомобилей) потоком воздуха, делятся на две категории: прямой и гибридный метод.

Прямой метод – уравнения Навье-Стокса решаются численно в области течения жидкости, включающей источник звука, а также часть ближнего акустического поля. При этом совместно определяются как параметры турбулентного течения, так и излучаемого акустического поля. Решение задачи в такой постановке получило название Direct Numerical Simulation – DNS. Однако данный метод крайне трудоемок, что не дает возможности применять его в полной мере при решении прикладных задач.

Гибридный метод основан на акустических аналогиях Лайтхилла. Расчет турбулентного потока жидкости, порождающего звук, производится в сравнительно небольшом конечном объеме. Далее по турбулентному полю вычисляется возникающий от этого поля звук (шум). В этой работе используется гибридный подход.

2. Уравнения акустики

В данной работе рассматривается создаваемый аэродинамическим путем звук, который является продуктом воздушного потока и отделен от звука, производимого вибрациями твердых тел. Воздушный поток может содержать флуктуации в результате неустойчивости, что дает при малых значениях чисел Рейнольдса правильную вихревую структуру. Шум, порожденный вихревой природой, называется вихревым. Звуком будем называть малые колебания сжимаемого газа.

Уравнения распространения звука получаются из основных уравнений динамики сжимаемого газа, которые включают в себя:

уравнение неразрывности –

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho v_i}{\partial x_i} = 0; \quad (2.1)$$

уравнения сохранения импульса (количества движения) –

$$\frac{\partial \rho v_i}{\partial t} + \frac{\partial}{\partial x_j} (T_{ij} + \rho v_i v_j) = 0, \quad (2.2)$$

где T_{ij} – тензор напряжений

$$T_{ij} = p\delta_{ij} + \mu \left(\frac{2}{3} \frac{\partial v_i}{\partial x_i} - \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \right). \quad (2.3)$$

В звуковых волнах любые необратимые процессы, включая вязкость и теплопроводность, должны приводить к увеличению полной энтропии и, соответственно, к нагреву среды, через которую проходит звуковая волна, и постепенной диссипации механической энергии звуковой волны.

Исследования показали, что эффекты, вызванные вязкостью и теплопроводностью газа, сводятся в основном к поглощению газа средой, которое обычно невелико. Поэтому при анализе процессов распространения звука в области частот интервала слышимости пренебрегают вязкостью и теплопроводностью, рассматривая процесс как адиабатический.

$$\frac{p}{\rho^\gamma} = const. \quad (2.4)$$

Пусть в среде с параметрами $p_0, \rho_0, v_0=0$ распространяется звук. Положим

$$p = p_0 + \tilde{p}, \quad \rho = \rho_0 + \tilde{\rho}, \quad v = \tilde{v}, \quad (2.5)$$

где волной обозначены малые величины, характеризующие звук. Таким образом, звуковые колебания определяются только «избыточными» давлением и плотностью. Подставляя равенства

уравнения (2.1), (2.2) (при $\mu=0$) и (2.4) и

пренебрегая слагаемыми второго порядка малости, получим уравнения акустики

$$\frac{\partial \tilde{\rho}}{\partial t} + \rho_0 \frac{\partial \tilde{v}_i}{\partial x_i} = 0; \quad (2.6)$$

$$\rho_0 \frac{\partial \tilde{v}_i}{\partial t} + \Delta \tilde{p} = 0; \quad (2.7)$$

$$\tilde{p} = c_0^2 \tilde{\rho}. \quad (2.8)$$

Здесь $c_0 = \sqrt{\gamma \frac{p_0}{\rho_0}}$ – скорость звука в невозмущенной среде. Отсюда нетрудно получить уравнение для возмущений плотности

$$\frac{\partial^2 \tilde{\rho}}{\partial t^2} - c_0^2 \Delta \tilde{\rho} = 0. \quad (2.9)$$

3. Акустическая аналогия Лайтхилла

Основы теории генерации звука потоком были заложены Лайтхиллом в работах [1, 2]. Пусть имеется некоторая пространственная область Ω , в которой происходит течение жидкости, описываемое уравнениями (2.1)– (2.3), а в остальном пространстве жидкость покоится и имеет параметры p_0, ρ_0 . Продифференцируем уравнение (2.1) по t , а (2.2) умножим на (-1) и продифференцируем по x_i . Складывая полученные равенства, будем иметь

$$\frac{\partial^2 \rho}{\partial t^2} - \frac{\partial^2}{\partial x_i \partial x_j} (T_{ij} + \rho v_i v_j) = 0.$$

Вычитая здесь из обеих частей $c_0^2 \Delta \rho$, получим:

$$\frac{\partial^2 \rho}{\partial t^2} - c_0^2 \Delta \rho = \frac{\partial^2 \tilde{T}_{ij}}{\partial x_i \partial x_j}, \quad (3.1)$$

где

$$\tilde{T}_{ij} = T_{ij} + \rho v_i v_j - c_0^2 \rho \delta_{ij} = (p - c_0^2 \rho) \delta_{ij} + \rho v_i v_j + \mu \left(\frac{2}{3} \frac{\partial v_i}{\partial x_i} - \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \right). \quad (3.2)$$

– так называемый тензор Лайтхилла.

Сравнивая уравнения (2.9) и (3.1), можно сформулировать аналогию Лайтхилла: флуктуации плотности в реальной сжимаемой жидкости, находящейся в произвольном движении, совпадают с флуктуациями плотности в однородной покоящейся среде, находящейся под действием внешнего силового поля (выражаемого правой частью уравнения (3.1)). Другими словами, можно считать, что в области Ω генерируется звук источниками, определяемыми правой частью уравнения (3.1), и распространяется вовне Ω .

Далее преобразуем выражение (3.2) для \tilde{T}_{ij} , учитывая, что в процессах распространения звука пренебрегаем вязкостью и теплопроводностью. Полагая $\mu=0$ и принимая во внимание (2.8), получим $\tilde{T}_{ij} = \rho v_i v_j$. Так как в адиабатическом (изэнтропическом) течении $\rho = \rho_0 + O(M^2)$, то при малых числах Маха можно принять $\tilde{T}_{ij} = \rho_0 v_i v_j$. Окончательно уравнение (3.1) запишется в виде

$$\frac{\partial^2 \rho}{\partial t^2} - c_0^2 \Delta \rho = \rho_0 \frac{\partial^2 v_i v_j}{\partial x_i \partial x_j}. \quad (3.3)$$

Так как уравнение (3.3) теперь рассматривается как уравнение распространения звука (флуктуаций плотности), то ρ в левой части следует понимать как $\tilde{\rho} = \rho - \rho_0$. Звуковое давление в соответствии с (2.8) будет вычисляться по формуле

$$p - p_0 = c_0^2 (\rho - \rho_0). \quad (3.4)$$

4. Влияние твердых границ на генерацию звука потоком

Рассмотрим теперь, что произойдет, если в потоке имеются твердые границы. Можно ожидать, что роль твердых границ проявится по крайней мере в трех отношениях.

Во-первых, звук, генерируемый квадрупольями, будет отражаться от границ и испытывать на них дифракцию.

Во-вторых, поток жидкости может вызывать появление соответствующим образом распределенных на самих границах источников типа монополей и диполей.

В-третьих, благодаря тем или иным твердым телам, находящимся в потоке, при малых числах Рейнольдца возможна обратная реакция процесса, связанного с порождением звука, на сам поток, вследствие чего возникают автоколебания.

Используя общие методы решения неоднородного волнового уравнения, на основе теории Лайтхилла результат получил Керль [3]. Он исходил из общей формулы Кирхгофа, дающей решение волнового уравнения (3.1) для акустического давления в точке наблюдения в виде

$$\rho(\mathbf{x}, t) = \frac{1}{4\pi} \int_S \left\{ \frac{1}{r} \left[\frac{\partial \rho}{\partial n} \right] + \frac{1}{r^2} \frac{\partial r}{\partial n} [\rho] + \frac{1}{c_0 r} \frac{\partial r}{\partial n} \left[\frac{\partial \rho}{\partial t} \right] \right\} dy + \frac{1}{4\pi c_0^2} \int_{\Omega} \frac{\partial^2}{\partial y_i \partial y_j} \left[\frac{T_{ij}}{r} \right] dy. \quad (4.1)$$

Здесь S – имеющаяся в потоке поверхность, n_i – составляющие вектора нормали, внешней по отношению к жидкости. В результате тождественных преобразований Керль получил следующую более удобную формулу:

$$\tilde{\rho}(\mathbf{x}, t) = \frac{1}{4\pi c_0^2} \int_S \frac{\partial}{\partial t} [\rho v_i n_i] \frac{dy}{r} - \frac{1}{4\pi c_0^2} \frac{\partial}{\partial x_i} \int_S [(\rho v_i v_j + T_{ij}) n_j] \frac{dy}{r} + \frac{1}{4\pi c_0^2} \frac{\partial^2}{\partial x_i \partial x_j} \int_\Omega [\tilde{T}_{ij}] \frac{dy}{r}. \quad (4.2)$$

Здесь T_{ij} – тензор напряжений (2.3), \tilde{T}_{ij} – тензор Лайтхилла (3.2). Смысл

этого выражения заключается в том, что общее излучение состоит из трех частей [4]. Первый интеграл представляет собой акустическое поле простых источников (монополей), распределенных на поверхности S . Второй интеграл представляет собой излучение диполей, распределенных на S . Эти источники связаны с пульсациями давления и вязких напряжений. При этом он включает в себя такие эффекты как отражение и диффракция звука на твердых поверхностях. Третий интеграл дает квадрупольное излучение в объеме Ω .

Для монополя, диполя и квадрупольного излучения интенсивность при малых числах Маха, соответственно, пропорциональна M , M^3 , M^5 . Таким образом, самым эффективным излучателем является монополь, затем диполь, и наименее эффективен квадруполь.

Для акустически жестких поверхностей ($v_i n_i = 0$) монополи отсутствуют, и формула упрощается. Как отмечено в [5], хотя вблизи твердых поверхностей влияние вязкости и теплопроводности газа возрастает, оно все же остается незначительным, и им по-прежнему можно пренебречь. Тогда формула **Ошибка! Источник ссылки не найден.** упрощается:

$$\tilde{\rho}(\mathbf{x}, t) = -\frac{1}{4\pi c_0^2} \frac{\partial}{\partial x_i} \int_S [p] n_i \frac{dy}{r} + \frac{1}{4\pi c_0^2} \frac{\partial^2}{\partial x_i \partial x_j} \int_\Omega [\rho_0 v_i v_j] \frac{dy}{r}. \quad (4.3)$$

Здесь также предполагается, что число M мало.

5. Основные характеристики звука

Характеристиками звука, которые могут быть восприняты человеческим ухом или измерены приборами, являются его интенсивность и спектральная плотность.

Источники шума	Расстояние в м	Уровень громкости в фонах
1. Шум скоростного самолёта с поршневым мотором . . .	5 м	130
2. Клёпка котла (внутри) . . .	от вента	113
3. Шум в поезде метро на большой скорости	—	90—95
4. Фортиссимо оркестра	—	90—100
5. Шумная улица	—	80—85
6. Громкий разговор	1	65—70
7. Машинописное бюро	—	75—80
8. Грузовой автомобиль	10—20	65—75
9. Шум в квартире (в городе)	—	40—50
10. Тихий разговор	1	40—50
11. Шумовой фон ночью (в городе)	—	38
12. Тиканье часов	0,5	30
13. Шопот	1,5	10
14. Шелест листьев при слабом ветре	—	10
15. Порог слышимости	—	0

Рис. 5.1. Порядки величин фоновых звуков

Интенсивность звука $I(\mathbf{x})$ представляет собой скорость, с которой звуковая энергия пересекает единичную площадку, перпендикулярную движению волны, в заданной точке \mathbf{x} . На достаточно большом расстоянии от источника интенсивность определяется выражением

$$I(\mathbf{x}) = \frac{1}{\rho_0 c_0} \text{Var}[\tilde{p}(\mathbf{x})] \quad (5.1)$$

и измеряется в $\text{Вт}/\text{м}^2$. Однако если речь идет о чувственном восприятии звука, то более подходящей является логарифмическая шкала уровней интенсивности, потому что для данной частоты в Гц ухо воспринимает одинаковые отличия в громкости при одинаковых отличиях логарифма интенсивности, а не ее самой. Уровень интенсивности по шкале децибелов (дБ) определяется как

$$IL(\mathbf{x}) = 10 \log \frac{I(\mathbf{x})}{I_{ref}},$$

где базовое значение интенсивности $I_{ref} = 10^{-12} \text{ Вт} \cdot \text{м}^{-2}$ (едва слышимый человеческим ухом звук на частоте 1000 Гц).

Типичный минимальный уровень интенсивности, при котором слышен звук, на типичных «высоких» частотах (от 500 до 8000 Гц) составляет 0 дБ, что соответствует $I = 10^{-12} \text{ Вт} \cdot \text{м}^{-2}$. Для понимания порядков величин фоновых звуков приведена табл. 5.1 [6].

Чтобы получить полную излучаемую акустическую мощность, необходимо проинтегрировать интенсивность по сфере, окружающей источник.

Другой используемой на практике величиной, является уровень звукового давления, распределенный по частотам, или полный уровень звукового давления (sound pressure level – SPL)

$$SPL(\mathbf{x}, f) = 10 \log \frac{S(f) \cdot \Delta f}{p_{ref}^2}, \quad SPL(\mathbf{x}) = 10 \log \frac{1}{p_{ref}^2} \text{Var}[\tilde{p}(\mathbf{x})] = 20 \log \frac{\tilde{p}_{rms}(\mathbf{x})}{p_{ref}}, \quad (5.2)$$

где базовое давление $p_{ref} = 2.04 \cdot 10^{-5} \text{ Па}$ приблизительно соответствует I_{ref} в плоской волне, а индекс *rms* (root-mean square) указывает на среднеквадратическое отклонение. Как уровень интенсивности, так и уровень звукового давления измеряются в децибелах.

6. Численные методы

Для получения акустических характеристик потока, обтекающего твердое тело, последовательно выполняем следующие действия.

- Гидродинамический расчет течения жидкости, который выдает гидродинамическое давление в узлах заданной сетки в любой момент времени.
- Вычисление акустических давлений по формуле

$$\tilde{p}(\mathbf{x}, t) = \frac{1}{4\pi c_0} \int_S \frac{(x_i - y_i) n_i}{r^2} \frac{\partial}{\partial t} [p] dy + \frac{1}{4\pi} \int_S \frac{(x_i - y_i) n_i}{r^3} [p] dy. \quad (6.1)$$

Используются численные методы вычисления интегралов и временных производных. Для вычисления поверхностных интегралов (в нашем случае на четырехугольной сетке) воспользуемся билинейной интерполяцией, а для временной производной воспользуемся конечно – разностной формулой третьего порядка.

- По полученному акустическому давлению рассчитывается спектр (используется метод сглаживания Бартлета) [7, 8].
- Получение характеристик звука (раздел 5).

7. Расчет зеркал

Имеются две формы (рис. 7.1) боковых зеркал дальнего вида. Необходимо получить звуковые характеристики для каждого из них. Для зеркала GMT360 был проведен

эксперимент по обдуванию его воздушным потоком с различными скоростями. С этими данными и будем сравнивать результаты наших расчетов.

Для расчета турбулентного потока за зеркалом использовался программный комплекс CFX с различными моделями турбулентного течения: URANS, DES SST, LES Smagorinsky. В ходе работы было сформулировано утверждение, что наиболее полно описывает процесс образования звуковых флуктуаций именно LES Smagorinsky. Все дальнейшие результаты приведены именно для этой схемы.

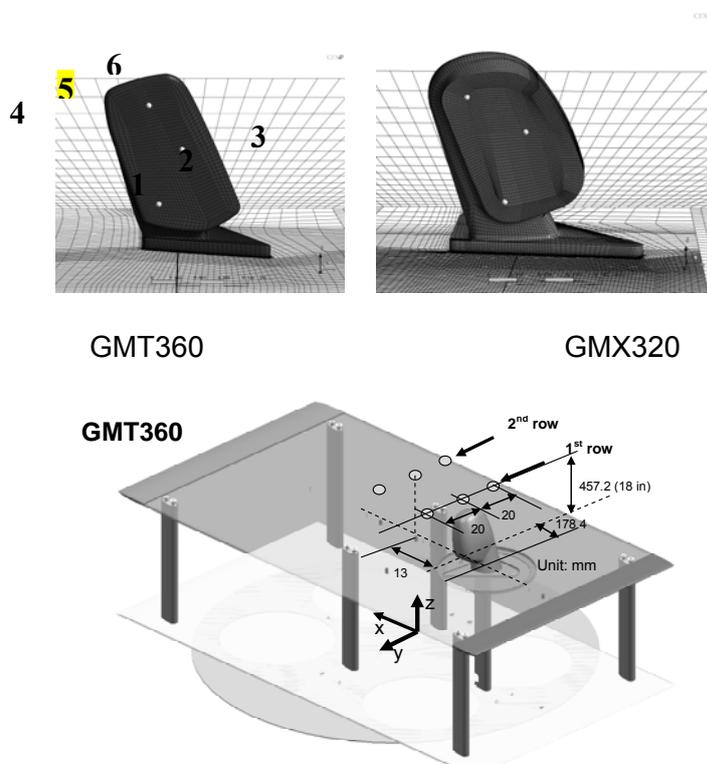


Рис. 7.1. Форма зеркал и экспериментальная установка

8. Экспериментальные данные и результаты расчетов

Предоставленные экспериментальные данные включали в себя частотное распределение уровня звукового давления (SPL) в 6 позициях микрофонов, расположенных на высоте 18 дюймов от пола (рис. 8.4).

Данные были представлены для одного зеркала (GMT360), но для разных скоростей набегающего потока воздуха: 30, 50, 70 и 90 миль в час (13.4, 22.35, 31.3 и 40.2 метров в секунду соответственно). Измеренные звуковые характеристики были даны как в присутствии модели зеркала, так и без нее (т.е. шум установки).

Для разделения звука от зеркала и фонового шума можно воспользоваться формулой:

$$SPL_{eff}(f) = 10 \lg \left(10^{\frac{SPL_2(f)}{10}} - 10^{\frac{SPL_1(f)}{10}} + 1 \right). \quad (8.1)$$

Расчетная модель не включает в себя учет внешнего шума (шума, производимого экспериментальной установкой), поэтому расчетные значения уровней звукового давления будем сравнивать с экспериментальным эффективным уровнем.

Разбиение микрофонов по группам сделано из-за различий в поведении звукового давления, зафиксированного ими. В зоне турбулентного течения преобладают низкие частоты, величиной до 2000Гц, порожденные находящимся в потоке зеркалом. Эффек-

тивная величина SPL не приносит качественных изменений в исходный экспериментальный сигнал, количественное уменьшение незначительно.

Микрофоны, вмонтированные в основание площадки, фиксируют псевдозвук, т.е. флуктуации гидродинамического давления, а также звук, порождаемый самим корпусом зеркала. Модуль гидродинамического давления значительно больше акустического, а если еще учесть случайный характер образования вихрей, то для расположенного в турбулентной зоне микрофона наблюдается более изрезанная кривая по сравнению с позицией 5, в которой имеется только акустическое давление. Кроме того, схемы со сглаживанием уменьшают амплитуду сигнала в пограничном слое, что в конечном итоге сказывается в виде более низкого уровня звукового давления (рис. 8.2, 8.3).

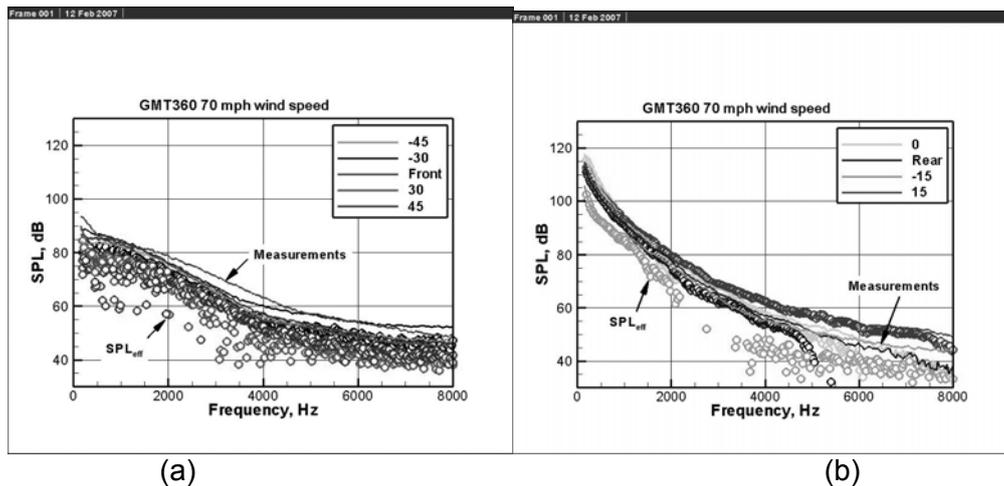
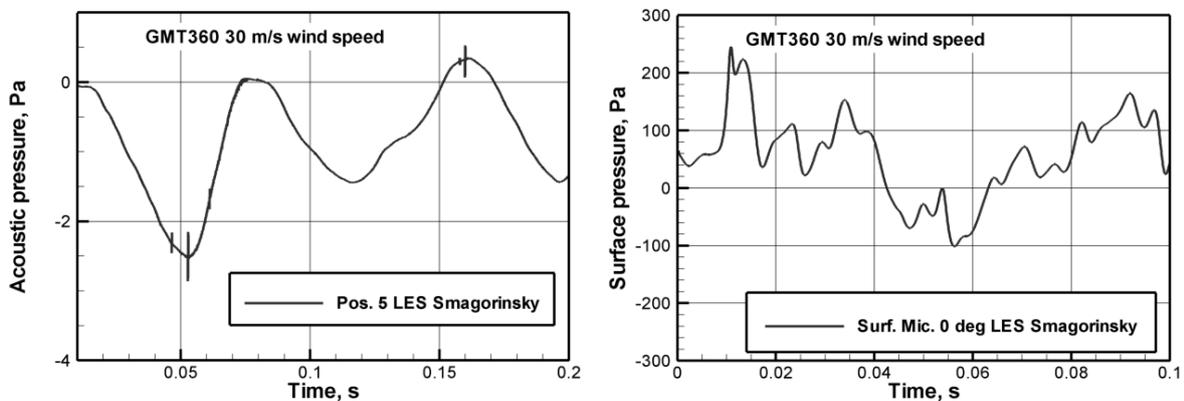


Рис. 8.1. Уровень звукового давления, измеренный в эксперименте микрофонами, вмонтированными в пол. (a) – Эффективный SPL вне турбулентной зоны за зеркалом. (b) – Эффективный SPL внутри турбулентной зоны за зеркалом.



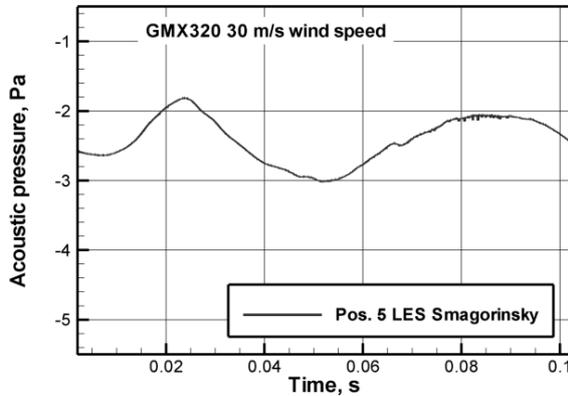
(a) Акустическое давление в позиции 5

(b) Давление на поверхности (высота 18 дюймов) от пола

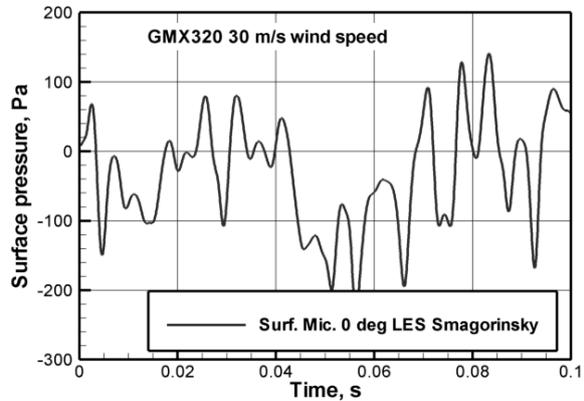
Рис. 8.2. Временное распределение давления: а) акустическое давление, рассчитанное акустическим постпроцессором по результатам CFX, полученным тремя моделями турбулентного течения; б) поверхностное давление рассчитанное CFX тремя моделями турбулентного течения (GMT360 модель зеркала, скорость потока 30 м/с)

Несколько раньше нами рассматривалось изменение акустического давления во времени, сейчас посмотрим на распределение в пространстве. Посмотрим на волновую природу акустического сигнала в дальнем (рис. 8.5) поле. Можно видеть периодичность в распространении звука, дипольность природы, его порождающей.

Направление распространения звука можно предсказать, только обладая определенным опытом, или определить в результате расчета. По распределению интенсивности звука (рис. 8.6) можно заключить о направлении излучения каждым из зеркал. На приведенных рисунках в каждой точке сферы единичного радиуса, с центром в начале координат, рассчитана интенсивность звуковой волны. В направлении излучения интенсивность больше.

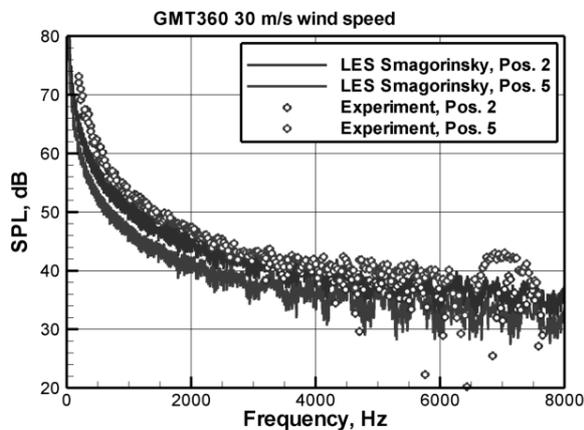


(a) Акустическое давление в позиции 5

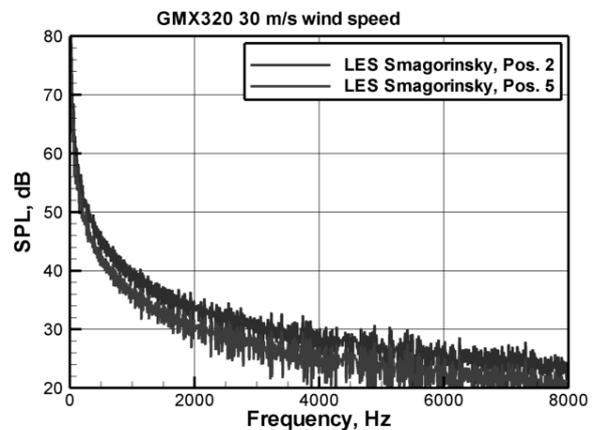


(b) Давление на поверхности (высота 18 дюймов)

Рис. 8.3. Временное распределение давления: а) акустическое давление рассчитанное акустическим постпроцессором по результатам CFX полученным тремя моделями турбулентного течения; б) поверхностное давление рассчитанное CFX тремя моделями турбулентного течения (GMX320 модель зеркала, скорость потока 30 м/с)



(a) GMT360



(b) GMX320

Рис. 8.4. Частотное распределение уровня звукового давления по двум позициям 2 и 5 на высоте 18 дюймов (данные позиции – центральные по отношению к набегающему потоку). Течение рассчитано комплексом CFX с тремя моделями турбулентного течения. Скорость набегающего потока 30 м/с. а) – GMT360 модель зеркала (также приведены результаты эксперимента). б) – GMX320 модель зеркала

Однако высчитать интенсивность звукового излучения по сфере – не единственный способ определить направление распространения звука. В этом также может помочь построение диаграммы направленности (рис. 8.6). Ее главный недостаток – она не может быть построена в пространстве, только в сечении, на плоскости. В этой работе она используется для подтверждения выводов, сделанных на основании распределения интенсивности

По рисункам можно судить о направлениях распространения звуковой энергии в пространстве – основная часть шума идет по направлению потока за зеркалом.

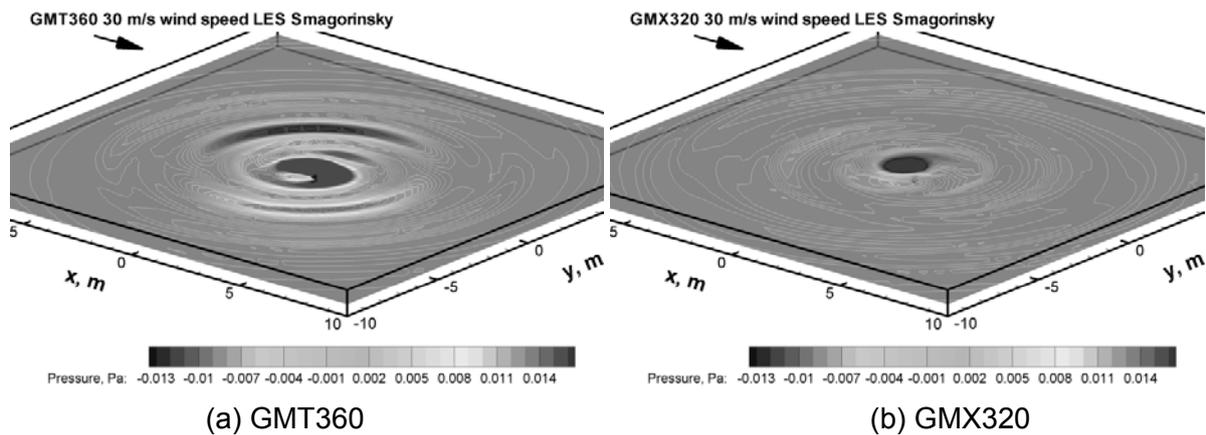


Рис. 8.5. Расчетное распределение звуковых давлений в дальнем поле (ведется учет излучения только от зеркала). Показана горизонтальная плоскость, расположенная на высоте 18 дюймов от поверхности пола. Направление ветра указано стрелкой. Расчет течения выполнен в CFX при скорости ветра 30 м/с

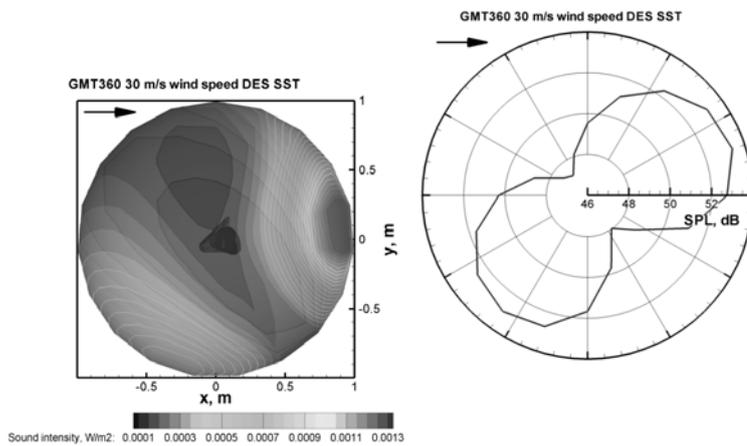


Рис. 8.6. Распределение звуковой энергии по сфере единичного радиуса (вид сверху) и диаграмма направленности (распределение полного уровня звукового давления (по окружности радиуса 1 метр на высоте 18 дюймов над уровнем пола) в полярной системе координат). Поток обсчитывался по схеме LES SM, скорость 30 м/с. Направление ветра указано на картинке стрелкой

модель зеркала	излучающая поверхность	вычисленная полная интенсивность (мощность), Вт
GMT360	только поверхность зеркала	0.054
	поверхность зеркала и основания	5.35
GMX320	только поверхность зеркала	0.026
	поверхность зеркала и основания	2.65

Таблица 8.1. Полная суммарная интенсивность звука, произведенного поверхностью зеркала и поверхностью стола. Интегрирование производилось по сфере единичного радиуса с центром в начале координат. Расчет произведен по данным CFX для DES SST-модели турбулентности при скорости ветра в 30 м/с

Помимо направления распространения звука, практический интерес представляет также мощность излучения, но эту величину затруднительно получить экспериментально. Мощность (суммарная интенсивность) может выступать в качестве критерия для сравнения излучающих свойств каждого из зеркал. Для получения этой величины проинтегрируем интенсивность звука, порождаемого зеркалом и полом, по ограничивающей поверхности. Результаты приведены в табл. 8.1.

На основании данных результатов можно заключить, что, во-первых, 320-е зеркало обладает лучшими акустическими свойствами. Во-вторых, важно учитывать влияние поверхности пола на образование звука. Обладая большой площадью, к тому же, содержащее зону вихревого потока, основание дает существенный вклад в излучаемую мощность.

9. Заключение

Разработан инструмент, позволяющий определять самые разнообразные характеристики шума. Реализовано программное обеспечение с дружественным пользовательским интерфейсом и модулем визуализации полученных результатов. Проведена проверка программы на тестовой задаче – определение характеристик звука, излучаемого при колебаниях цилиндра в потоке воздуха. Определены причины возникновения звука. Исследованы методы нахождения состоятельной оценки спектральной плотности мощности звукового излучения.

Проведен сравнительный анализ гидродинамических моделей расчета турбулентного течения и их влияния на генерацию звука. Наиболее предпочтительной с точки зрения детализации акустического давления и количественного совпадения с экспериментом оказалась LES Smagorinsky модель. С помощью разработанного инструмента были найдены уровни звукового давления, возникающего при обтекании двух автомобильных зеркал, проведено сравнение расчетных данных с экспериментом.

Помимо этого, путем математического моделирования были получены данные, которые затруднительно измерить экспериментально: поля акустических давлений для зеркал, по которым было сделано заключение о дипольной природе возникающего звука; интенсивность излучения, по которой появилась возможность указать направление распространения звука. Рассчитана суммарная мощность излучения – параметр, по которому можно сравнивать различные, излучающие звук объекты.

Сделаны подробные выводы об акустических свойствах обоих зеркал.

Литература

1. Lighthill M.J. On sound generated aerodynamically. I. General theory // Proc. Roy. Soc. (London). – A 211. – 1952. – P. 564–587.
2. Lighthill M.J. On sound generated aerodynamically. II. Turbulence as a source of sound. // Proc. Roy. Soc. (London). – A 222. – 1954. – P. 1–32.
3. Зарембо Л.К., Красильников В.А. Введение в нелинейную акустику. – М.: Наука, 1966. – 520 с.
4. Лайтхилл Дж. Волны в жидкостях. – М.: Наука, 1981. – 208 с.
5. Блохинцев Д.И. Акустика неоднородной движущейся среды. – М.: Наука, 1981. – 208 с.
6. Стретт Д.В. Теория звука. – М.: ГИТТЛ, 1955. – 475 с.
7. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. М.: Мир, 1982. – 428 с.
8. Дж. Бендат, А. Пирсол. Прикладной анализ случайных данных. – М.: Мир, 1989. – 540 с. (J.S. Bendat, A.G. Piersol. Random Data. Analysis and Measurement Procedures. Wiley, New York, 1986).

АНАЛИЗ ПОТОКА ИЗОБРАЖЕНИЙ НА ОСНОВЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ

М.А. Лукин, Р.С. Увашев (Вятский государственный университет)
Научный руководитель – к.т.н., доцент В.С. Ростовцев
(Вятский государственный университет)

В докладе описан подход к обнаружению и классификации объектов в потоке изображений на основе искусственных нейронных сетей с использованием распределенных вычислений, анализу их поведения и его прогнозированию.

Введение

Распознавание растровых изображений при помощи искусственных нейронных сетей (ИНС) является одной из наиболее распространенных задач искусственного интеллекта. Существует множество алгоритмов для повышения точности, помехоустойчивости и скорости обучения, а также программно-аппаратных средств распараллеливания вычисления значений выходов ИНС для получения результата за кратчайшие промежутки времени. Однако для изучения динамики объектов исследуемого пространства (ИП) в реальном масштабе времени требуются дополнительные вычислительные затраты и особый подход к распределению ресурсов между компонентами вычислительной системы (ВС). В разрабатываемом проекте авторы попытались обобщить известные методы распознавания и анализа потока изображений с использованием распределенных вычислений, чтобы отслеживать и прогнозировать поведение объектов исследуемого пространства в реальном масштабе времени.

1. Концепция распознавания объектов

Анализ и прогнозирование поведения объектов в реальном масштабе времени возможны при условии достаточно быстрого получения информации об их состоянии на текущий момент (координат обнаруженных объектов на поступившем к обработке видеокadre). При этом размер и положение объекта на изображении определяются относительно источника видеосигнала (ИВС). Для распознавания объектов можно выделить два способа:

- 1) разработка и обучение ИНС со сложной структурой, способной выделять на всем изображении множество заранее определенных объектов, представленных в любом масштабе и ракурсе;
- 2) разбиение всего изображения на подизображения и распознавание каждого на отдельном узле ВС.

Последний из рассмотренных подходов авторы считают более перспективным по ряду причин, среди которых:

- 1) простая структура ИНС, не решающая задачу выделения объекта на всем изображении, а лишь распознающая его целиком на предоставленном ей фрагменте;
- 2) сокращение времени обучения ИНС как следствие предыдущего пункта;
- 3) возможность программной реализации распределенных вычислений (аппаратные средства для передачи части исходного изображения для обработки другому узлу не являются обязательными, в то время как для реализации ИНС со сложной структурой может потребоваться нейрокомпьютер);
- 4) более эффективное распределение ресурсов ВС, локализация вычислений на отдельно взятом узле;
- 5) решение проблемы повторно встречающегося объекта.

Выделение подизображений передвижением окна сканирования по исходному изображению проиллюстрированы на рис. 1.

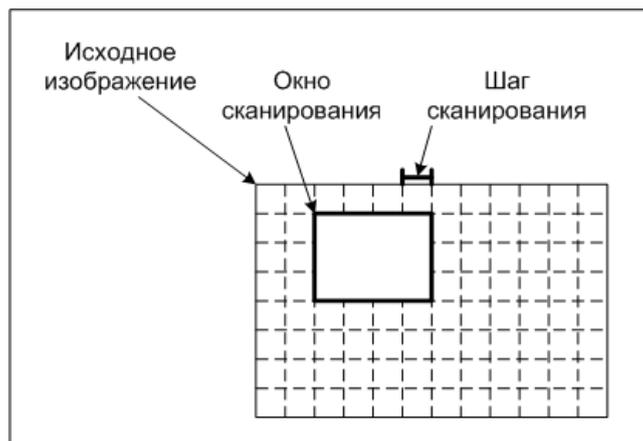


Рис. 1. Сканирование изображения

Поскольку число входов ИНС постоянно, распознаваемое изображение масштабируется одним из выбранных алгоритмов до заранее определенного размера. Если в результате распознавания ИНС удалось классифицировать объект, информация о нем (идентификатор, координаты, масштаб) может быть передана на соответствующий узел ВС для анализа его поведения. В противном случае изображение, аналогично своему предшественнику, разбивается на подизображения, которые затем распределяются по свободным узлам ВС.

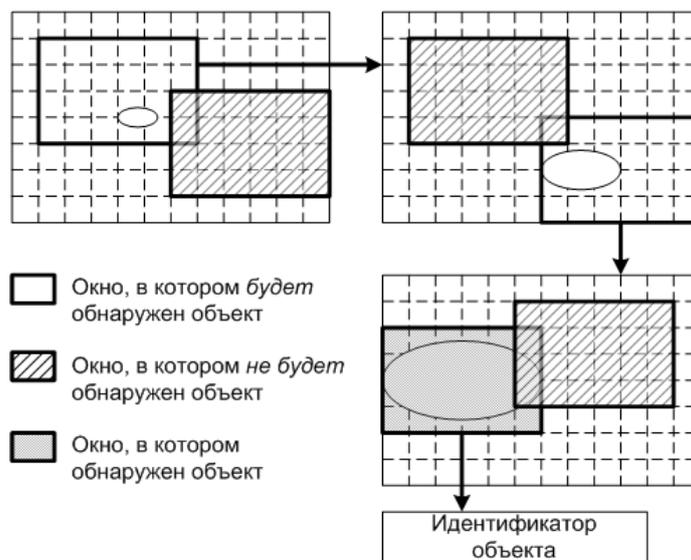


Рис. 2. Пример последовательности распознавания

Глубина такой рекурсии определяется скоростью обработки каждого фрагмента, количеством узлов ВС и частотой пополнения потока новыми изображениями. При недостаточном количестве ресурсов ВС может возникнуть «эффект близорукости» (глубина рекурсии слишком мала, и достаточно детальное изучение изображения становится невозможным), в худшем (если скорость поступления изображений выше скорости их обработки) становится невозможным анализ потока в реальном масштабе времени. Изображение, которое не удалось классифицировать как объект, разбивается на фрагменты при условии, что в очереди к данному узлу нет изображений с более высоким приоритетом (в простейшем случае приоритет определяется временем поступления исходного изображения на обработку). Таким образом, система постепенно «вглядывается» в исходное изображение, пока ее «внимание» не привлечет более новое. При выборе

слишком маленького шага сканирования может потребоваться значительное количество вычислительных узлов, в связи с чем оправдано использование вспомогательной ИНС для определения координат окна сканирования, содержащего «сгустки» информации.

Описанные выше действия проиллюстрированы на рис. 2.

2. Архитектура искусственной нейронной сети

После масштабирования для распознавания объекта на выделенном фрагменте можно применить известные типы ИНС в зависимости от конкретной задачи и качества изображений в потоке [2–4]. Ввиду простоты реализации и высокой скорости обучения на данный момент авторы отдают предпочтение сетям встречного распространения (СВР) [1]. В СВР объединены два хорошо известных алгоритма: самоорганизующаяся карта Кохонена и звезда Гроссберга. При этом появляются свойства, которых нет ни у одного из них в отдельности. СВР функционирует подобно столу справок, способному к обобщению. В процессе обучения входные векторы ассоциируются с соответствующими выходными векторами; они могут быть двоичными, состоящими из нулей и единиц, или непрерывными. Когда сеть обучена, приложение входного вектора приводит к требуемому выходному вектору. Обобщающая способность сети позволяет получать правильный выход даже при неполном или искаженном входном векторе. Сети такого типа обычно используются для распознавания и восстановления образов.

Схема СВР без обратных связей, которая может быть использована для классификации изображений, представлена на рис. 3.

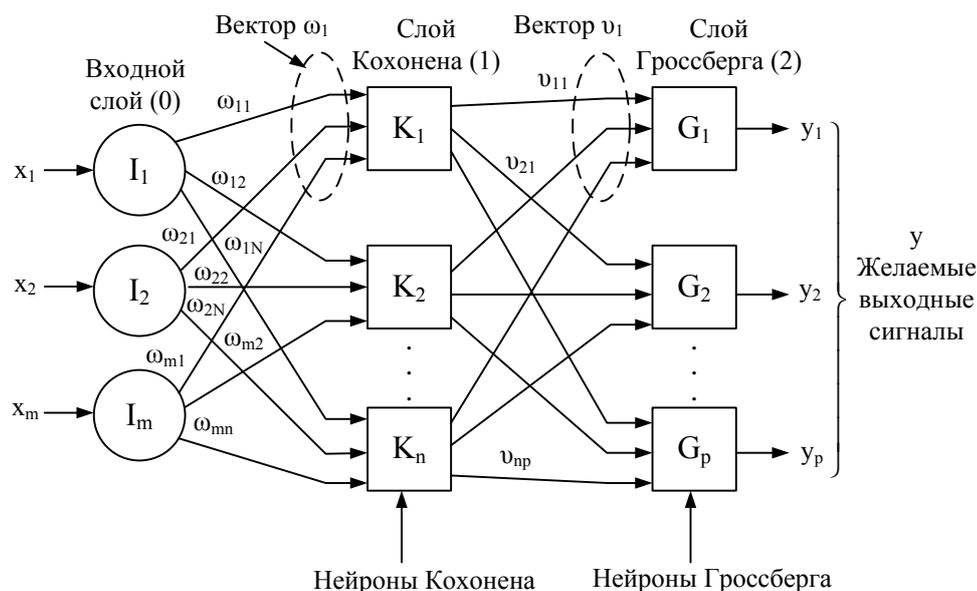


Рис. 3. Сеть встречного распространения

Нейроны слоя 0 (показанные кружками) служат лишь точками ветвления и не выполняют вычислений. Каждый нейрон слоя 0 соединен с каждым нейроном слоя 1 (называемого слоем Кохонена) отдельным весом u_{mn} . Аналогично, каждый нейрон в слое Кохонена (слое 1) соединен с каждым нейроном в слое Гроссберга (слое 2) весом u_{np} . СВР функционирует в двух режимах: в нормальном режиме, при котором принимается входной и выдается выходной вектор, и в режиме обучения, при котором подается входной вектор и веса корректируются, чтобы дать требуемый выходной вектор.

Для обучения сети, однако, в общем случае затруднительно сформировать обучающую выборку (ОБ) так, чтобы она содержала изображения всех объектов, которые предполагается обнаруживать в ИП. Поэтому результатом классификации является не

номер, а идентификатор объекта, представляющий собой совокупность «коэффициентов похожести» на объекты ОВ (например, если объектом классификации является автомобиль, то автомобиль марки, не выпущенной на момент формирования ОВ, может быть классифицирован как похожий на два или более автомобиля из тех, которые в ней уже присутствовали). Коэффициентами «похожести» в данном случае будут являться выходы СВР, а критерием удачной классификации изображения – превышение порога некоторой функцией, которая в простейшем случае будет вычисляться как среднеквадратичное значение этих коэффициентов:

$$F^2 = \sum_{i=1}^p y_i^2 > F_{\min}^2, \quad (1)$$

где p – количество нейронов в выходном слое, y_i – значение выхода i -го нейрона выходного слоя, F – значение функции «похожести», F_{\min} – пороговое значение функции «похожести». Таким образом, ОВ не содержит весь набор распознаваемых объектов, а служит лишь базисом для его определения, и объекты с близкими по значению идентификаторами на разных видеокдрах будут интерпретироваться как один и тот же.

3. Анализ и прогнозирование

Под анализом поведения объектов в данном случае авторы понимают выявление следующих ситуаций:

- 1) появление нового объекта (на текущем видеокдре обнаружен объект, не присутствовавший на предыдущем);
- 2) исчезновение объекта (ситуация, обратная п. 1);
- 3) изменение положения объекта (различаются координаты объекта на предыдущем и текущем видеокдрах).

Ситуации (1) и (2) не подлежат дальнейшей обработке, в то время как информация об изменении координат объекта используется как для выявления критических ситуаций (например, объект переместился в некоторую «опасную» зону), так и для аппроксимации траектории его движения (определение вероятности возникновения критической ситуации). Для каждого объекта находится зависимость его координат в ИП от времени с использованием регрессионного анализа и метода наименьших квадратов [4]. При этом для «забывания» «старых» точек каждая сумма перед добавлением в нее новой компоненты последовательности умножается на коэффициент «затухания», в соответствии с формулами (2).

$$\begin{aligned} S_t &= \sum_{i=1}^T t_i \cdot k^{T-i}, \\ S_x &= \sum_{i=1}^T x_i \cdot k^{T-i}, \\ S_{xt} &= \sum_{i=1}^T x_i \cdot t_i \cdot (k^{T-i})^2, \\ S_{t^2} &= \sum_{i=1}^T t_i^2 \cdot (k^{T-i})^2, \\ S_n &= \sum_{i=1}^T k^{T-1}, \end{aligned} \quad (2)$$

где $S_t, S_x, S_{xt}, S_t^2, S_n$ – значения элементов матрицы для вычисления коэффициентов линейной регрессии, T – количество точек, t_i – значение времени в i -й точке, x_i – значение координаты в i -й точке, k – коэффициент «затухания».

Значение коэффициента «затухания» выбирается в интервале $(0; 1]$ и чем ближе к нулю, тем менее «старые» точки влияют на прогнозирование движения объекта. Коэффициенты регрессии вычисляются методом Крамера по формуле (3):

$$\begin{aligned} \Delta &= S_n \cdot S_t^2 - S_t^2, \\ a &= \frac{S_x \cdot S_t^2 - S_{xt} \cdot S_t}{\Delta}, \\ b &= \frac{S_n \cdot S_{xt} - S_t \cdot S_x}{\Delta}, \end{aligned} \quad (3)$$

где Δ – детерминант матрицы, a, b – коэффициенты линейной регрессии.

Предполагаемое значение координаты от времени может быть вычислено по формуле:

$$x(t) = a + b \cdot t, \quad (4)$$

Теперь прогнозирование, (например, столкновения двух объектов) может быть выполнено путем вычисления значения пространственно-временного интервала на некотором промежутке времени и сравнения его с некоторой пороговой величиной, которую можно считать опасной в данной ситуации:

$$S^2(t) = (x_1(t) - x_2(t))^2 + (y_1(t) - y_2(t))^2 + (z_1(t) - z_2(t))^2 < \Delta_{on}, \quad (5)$$

где $S(t)$ – расстояние между объектами, $x_i(t), y_i(t), z_i(t)$ – координаты i -го объекта в момент времени t , Δ_{on} – критическое значение пространственно-временного интервала.

В общем случае для прогнозирования можно применить другие методы: полиномиальная аппроксимация, регрессионный анализ на рекуррентных ИНС и т.д.

4. Практическое использование

В настоящее время авторами ведется разработка программной реализации (ПР) описанного выше подхода на языке Java (<http://java.sun.com>) в интегрированной среде разработки netbeans (<http://www.netbeans.org>).

Для передачи сообщений между узлами ВС используется протокол UDP. Датаграмные сокет не гарантируют доставку пакетов данных, однако работают быстрее потоковых и обеспечивают возможность широковещательной рассылки пакетов данных одновременно всем узлам ВС [5].

Для передачи изображений используется протокол TCP, гарантирующий доставку пакетов данных и сохранение их целостности.

ПР состоит из двух частей: клиентской и серверной. Серверная часть функционирует на одном узле ВС и занимается распределением изображений потока между клиентами, а также обработкой результатов распознавания: анализом и прогнозированием поведения объектов. Клиентская часть работает на всех остальных узлах, ее роль заключается в приеме, сканировании и классификации изображения или его фрагмента, отправке результатов распознавания на сервер.

Такая реализация допускает динамическую конфигурацию ВС: добавление дополнительного узла в ВС приводит к немедленному включению его в работу и увеличению общей производительности ВС. Отказ любого узла ВС (за исключением серверной части) не приведет к потере работоспособности всей ВС, а лишь к потере информации о том фрагменте видеокadra, который обрабатывался на этом узле в момент отката, и снижению общей производительности ВС.

Заключение

В докладе рассмотрен метод обнаружения и распознавания объектов в потоке изображений в реальном масштабе времени с использованием распределенных вычислений. Выбрана архитектура ИНС и алгоритм ее обучения. Определен метод анализа и прогнозирования поведения объектов и приведено описание программной реализации выбранного подхода.

Литература

1. INTUIT.ru: Интернет-Университет Информационных Технологий – дистанционное образование [Электронный ресурс]. Г.Э. Яхъяева. Учебный курс – Основы теории нейронных сетей. – Режим доступа: <http://www.intuit.ru/department/ds/neuronnets/>
2. Саймон Хайкин. Нейронные сети: полный курс. 2-е издание.: Пер. с англ. – М.: Издательский дом «Вильямс», 2006. – 1104 с.: ил. – Парал. тит. англ.
3. Осовский С. Нейронные сети для обработки информации / Пер. с польского И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.: ил.
4. Выгодский М.Я. Справочник по высшей математике. – М.: ООО «Издательство Астрель»: ООО «Издательство АСТ», 2002. – 992 с.: ил.
5. Sun Microsystems – Россия и СНГ – Программирование на Java: подробное руководство [Электронный ресурс]. Александр Фролов. – Режим доступа: <http://ru.sun.com/java/books/online/index.html>

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ МАТЕМАТИЧЕСКОЙ МОДЕЛИ РАЗВИВАЮЩЕЙСЯ ЭКОНОМИКИ

А.В. Козлицкая, А.В. Кощев, А.И. Фетинина
(Вятский государственный университет)

Научный руководитель – к.ф.-м.н., доцент Н.Н. Оленев
(Вятский государственный университет)

Описывается создание математической модели межотраслевого баланса развивающейся экономики и идентификация параметров этой модели. Множества реальных отраслей экономики агрегировано в три сектора. Приведены программные коды идентификации параметров матрицы производственных затрат.

Введение

В современном мире трудно представить жизнь без использования компьютерной техники. Жизнь все время ускоряется и усложняется, идет постоянное развитие всех ее сфер и областей. Если раньше каждый показатель производительности компьютера удваивался раз в два года, то в последние годы этот период уменьшился до 18 месяцев. В связи с развитием многопроцессорных и многоядерных архитектур вычислительных систем параллельное программирование начинают применять повсеместно во всех областях науки, техники и бизнеса.

Параллельное программирование необходимо применять для решения таких задач, которые другим способом решить либо невозможно, либо чрезвычайно трудно. Это расширяет границы области научных исследований и приложений в бизнесе, в которых эффективно применять методы математического моделирования.

Моделирование развивающейся экономики – довольно сложный и трудоемкий процесс, так как развитие всегда сопровождается различными скачками, неопределенностью параметров модели. Поэтому с помощью статистических данных далеко не всегда можно сделать качественный анализ и прогноз будущих значений. С развитием высокопроизводительных систем эта задача может быть значительно упрощена и сведена к минимуму затрат. Благодаря этому создание математических моделей в наши дни осуществляется, в основном, с использованием различных компьютерных технологий [1].

Применение параллельных вычислений в моделировании экономики расширяет возможности разработчиков моделей. Появилась возможность идентифицировать внешние параметры сложных нормативных балансовых динамических моделей экономических систем в различных регионах и странах. Для этого разрабатываются специальные критерии близости и похожести для статистических и рассчитанных по модели временных рядов макропоказателей изучаемой экономической системы страны или региона. Кроме этого, применяются информационные технологии создания математических моделей с помощью интеллектуальных систем поддержки математического моделирования типа системы ЭКОМОД [2, 3], что позволяет уменьшить число независимых внешних параметров. Также совершенствуются методы глобальной оптимизации, которые сокращают время расчета.

Наша задача заключалась в создании имитационной модели экономики, параметры которой необходимо было идентифицировать на основе известных векторов валового выпуска и конечного потребления. С помощью найденных параметров стало возможным прогнозировать значения валового выпуска продукции по отраслям.

Идентификация параметров матрицы производственных затрат

Идентификация сложной математической модели той или иной экономической системы заключается в определении внешних параметров этой модели. В подобных за-

дачах параллельные вычисления позволяют находить те решения, которые без применения технологии распараллеливания было невозможно найти [1, 4].

До появления высокопроизводительных параллельных вычислений огромный возможный набор сочетаний значений параметров не позволял с приемлемой точностью определять эти значения. Но в настоящее время благодаря достаточно полному перебору становится возможным решить задачу идентификации параметров сложных математических моделей экономических систем.

Простейшая имитационная модель развивающейся экономики может быть представлена в виде трехсекторной нормативной балансовой динамической математической модели [5]. В такой модели необходимо идентифицировать всего девять параметров, что является наиболее приемлемым. Поэтому такую модель целесообразно было применить для поставленной перед нами задачи. Расчеты по модели дают количественную оценку динамики макропоказателей экономики как региона, так и страны в целом. Динамика материальных и финансовых балансов в модели задана изменением запасов продуктов, факторов производства и денег. Выделены следующие сектора экономики: инновационный сектор, сырьевой сектор и промышленность и прочие отрасли.

Наша модель построена на основе статистических данных макроэкономических показателей России за 1985–1990 годы. Первоначальные данные были разделены на семнадцать отраслей. Такое количество предполагает идентификацию двухсот восьмидесяти девяти параметров, вычисление которых является затруднительным. Поэтому рационально было объединить имеющиеся данные в три различных сектора экономики.

В инновационный сектор были включены четыре отрасли: электроэнергетика, нефтегазовая, химическая и нефтехимическая, пищевая промышленность. К сырьевому сектору были отнесены угольная и прочая топливная промышленность, черная и цветная металлургия, лесная, деревообрабатывающая и целлюлозная промышленность, сельское и лесное хозяйство. И в третий сектор вошли все оставшиеся отрасли: машиностроение и металлообработка, промышленность строительных материалов, легкая промышленность, грузовой транспорт и производственная связь, торговля и общественное питание, заготовки, материально-техническое снабжение и сбыт, прочие отрасли.

Все семнадцать отраслей изначально представлены в виде матрицы межотраслевого баланса Леонтьева [6]. Поэтому агрегирование такой матрицы в трехсекторную предполагает выполнение определенных операций.

Экономико-математическая модель межотраслевого баланса (МОБ) отражает производство и распределение ВВП по отраслям, межотраслевые потоки, использование материальных и трудовых ресурсов, создание и распределение национального дохода (вектора конечного потребления).

Экономико-математическая модель МОБ представляет собой систему уравнений, отражающих взаимосвязь включенных в его систему элементов. В матричном виде эта модель может быть представлена как

$$\vec{X} = A\vec{X} + \vec{Y}, \quad (1)$$

где $\vec{X} = \{X_1, X_2, X_3\}$ – вектор валовой продукции, $\vec{Y} = \{Y_1, Y_2, Y_3\}$ – вектор конечного потребления, $A = \{a_{ij}\}$ – матрица затрат, а элементы матрицы a_{ij} называют коэффициентами прямых затрат – это затраты i -ой отрасли на единицу валовой продукции j -ой отрасли.

Таким образом, в модели межотраслевого баланса выделяют валовую продукцию и конечную продукцию. Под валовым продуктом отрасли понимают весь объем ее продукции, произведенный в течение рассматриваемого периода. Конечный продукт отрасли представляет собой ту часть ее валового продукта, которая не используется в производстве в течение рассматриваемого периода как в самой отрасли, так и во всех остальных отраслях.

Матричное уравнение (1) можно записать в виде

$$\bar{X} - A\bar{X} = \bar{Y} \Rightarrow (E - A)\bar{X} = \bar{Y}. \quad (2)$$

Матричное уравнение (2) можно рассматривать для решения следующих задач:

1. определение объемов конечной продукции $\bar{Y} = \{Y_1, Y_2, Y_3\}$ по заданным объемам валовой продукции $\bar{X} = \{X_1, X_2, X_3\}$;
2. определение объемов валовой продукции отраслей $\bar{X} = \{X_1, X_2, X_3\}$ по заданным объемам конечной продукции $\bar{Y} = \{Y_1, Y_2, Y_3\}$.

Матрицу прямых затрат можно агрегировать. Агрегированием матрицы прямых затрат называется преобразование исходной матрицы $A = \{a_{ij}\}$ размерности $n \times n$ в матрицу прямых затрат меньшей размерности путем разбиения исходных i отраслей на конечное (меньшее) число I_k непересекающихся подмножеств (групп отраслей), называемых агрегатами, состоящими из I_l отраслей. При этом не должны нарушаться балансы производства и распределения продукта между отраслями.

Наша исходная матрица представляет собой матрицу $A = \{a_{ij}\}$ прямых затрат размерности (17×17) за 1990 г. Также задан вектор валовых выпусков $\bar{X} = \{X_1, X_2, \dots, X_{17}\}$.

Требуется построить агрегированную матрицу для трех укрупненных отраслей.

Коэффициенты этой матрицы $B = \{b_{ij}\}_{i,j=1..3}$ вычисляются следующим образом:

$$b_{kl} = \frac{\sum_{i \in I_k} \sum_{j \in I_l} a_{ij} x_j}{\sum_{j \in I_l} x_j}. \quad (3)$$

Реализация модели

В результате проделанной работы с использованием формулы (3) была получена агрегированная матрица за 1990 г. (таблица):

Инновационный сектор	0,2673	0,0726	0,0715
Сырьевой сектор	0,3301	0,269	0,093
Промышленность и прочие отрасли	0,1733	0,1406	0,3298

Таблица. Агрегированная матрица производственных затрат

По данной матрице были определены интервалы (сетка) изменения идентифицируемых параметров. Саму полученную матрицу нельзя использовать для составления прогнозов, поскольку она составлена по данным всего лишь одного года без учета предыдущих значений. Кроме этого, 1990 г. является одним из переломных годов в российской экономике. Поэтому для идентификации параметров данной матрицы и использования ее в дальнейшем для прогнозов была написана параллельная программа. Исходными данными для программы являются, во-первых, вектора валового и конечного выпусков за несколько лет (минимум – 5), а во-вторых – начальные и конечные значения по каждому из идентифицируемых параметров, в пределах которых осуществляется идентификация. В соответствии с теорией Леонтьева, матрица производственных затрат удовлетворяет условиям Фробениуса и Хоукинса-Саймона [6], поэтому все параметры – коэффициенты матрицы затрат – могут находиться в пределах от 0 до 1. Но в нашем случае для упрощения задачи интервалы были заданы по агрегированной матрице 1990 г. в виде «значение +/- значение, деленное пополам». Эти интервалы рас-

считываются непосредственно в программе и представлены в виде двух матриц – нижних и верхних границ интервалов.

Программа написана с использованием функций MPI в C++. Сначала была написана последовательная программа, а далее она была распараллелена на доступное число процессоров. Основная процедура, в которой непосредственно осуществляется распараллеливание и идентифицируются параметры матрицы производственных затрат, представлена ниже.

```
int main(int argc, char *argv[])
{
    float W[YEARS*OTRASL];
    float X[YEARS*OTRASL];
    float limits[OTRASL*OTRASL];
    float limits1[OTRASL*OTRASL];
    float limits2[OTRASL*OTRASL];
    float MY_X[YEARS*OTRASL];
    float MY_A[OTRASL*OTRASL];

    int numproc; // число процессоров
    int procind; // номер процессора

    MPI_Status status;
    MPI_Init(&argc, &argv);
    MPI_Comm_size(MPI_COMM_WORLD, &numproc);
    MPI_Comm_rank(MPI_COMM_WORLD, &procind);

    char perProcLogPath[20];
    sprintf(perProcLogPath, "./log/%d.log", procind);
    perProcLog = fopen(perProcLogPath, "w");
    clock_t start, finish;
    double duration;
    start = clock();

    float bestK = 0.0;
    float bestA[OTRASL*OTRASL];
    float tmp[1];
    readW("./data/acc.txt", tmp);
    float acc = tmp[0]; // точность
    readW("./data/w.txt", W);
    readW("./data/x.txt", X);
    readW("./data/a.txt", limits);

    // границы поиска параметров матрицы A
    for (int i = 0; i < OTRASL * OTRASL; i++) {
        limits1[i] = limits[i] / 2.0;
        limits2[i] = limits[i] * 1.5;
    }
    if (procind == 0) {
        printf("Accuracy: %.6f\n", acc);
        printf("W:\n");
        printW(W, YEARS, OTRASL, stdout);
        printf("X:\n");
    }
}
```



```

    }
    }
// bestK, bestA - лучший результат, он посылается главному процессу

if (procind == 0) {
    fprintf(stdout, "%d %.12f\n", procind, bestK);
    fflush(stdout);
    int bestN = 0;
    for (int tmp = 1; tmp < numproc; tmp++) {
        float bestK_;

        MPI_Recv(&bestK_, 1, MPI_FLOAT, tmp, 77, MPI_COMM_WORLD, &status);
        MPI_Recv(MY_A, OTRASL*OTRASL, MPI_FLOAT, tmp, 77,
MPI_COMM_WORLD, &status);

        fprintf(stdout, "%d %.12f\n", tmp, bestK_);
        fflush(stdout);
        if (bestK_ > bestK) {
            bestK = bestK_;
            bestN = tmp;
            copyArray(MY_A, bestA, OTRASL * OTRASL);
        }
    }
    fprintf(stdout, "Best K = %.12f [%d]\n", bestK, bestN);
    fflush(stdout);
    fprintf(stdout, "Best A:\n");
    printW(bestA, OTRASL, OTRASL, stdout);
    finish = clock();
    duration = (double)(finish - start) / CLOCKS_PER_SEC;
    printf( "%2.1f seconds\n", duration );
}
else {
    MPI_Send(&bestK, 1, MPI_FLOAT, 0, 77, MPI_COMM_WORLD);
    MPI_Send(bestA, OTRASL*OTRASL, MPI_FLOAT, 0, 77, MPI_COMM_WORLD);
}
fclose(perProcLog);
MPI_Finalize();
return 0;
}

```

В этой программе для первого года берутся статистические значения валового выпуска и конечного потребления. Кроме этого, в программе задается точность, с которой рассчитывается матрица производственных затрат. Для каждого последующего года с использованием идентифицированной матрицы производственных затрат вычисляются расчетные значения валового продукта

$$X(t+1) = AX(t) + Y(t+1). \quad (4)$$

После расчета векторов валовых продуктов по формуле (4) производится сравнение выходных временных рядов переменных модели (X) с доступными статистическими временными рядами (Y) за 1985–1990 гг. При этом используется индекс Тейла $E(X, Y)$. Но для удобства расчетов вместо индекса Тейла лучше применить коэффициент близости $U(X, Y) = 1 - E(X, Y)$. Чем выше он, т.е. чем ближе он к единице, тем более близки ряды.

$$U(X, Y) = 1 - 3 \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{\sum_{i=1}^n X_i^2 + \sum_{i=1}^n Y_i^2}}. \quad (5)$$

На каждом процессоре вычисляются матрицы производственных затрат, векторы валового выпуска и индексы близости (формула (5)). Вариант с лучшим значением коэффициента близости отправляется главному процессору, который выбирает лучший из лучших вариантов. Результаты – лучший коэффициент близости, лучшая матрица производственных затрат и время счета – выводятся главным процессором на экран.

Заключение

С помощью данной программы были идентифицированы параметры модели развивающейся российской экономики. Но, тем не менее, данная модель является универсальной (она применима на любом количестве процессоров и с некоторой модификацией – для большего числа отраслей).

С помощью модели можно, задавая начальные значения по валовому и конечному продукту, рассчитывать матрицу производственных затрат, а уже далее находить прогнозные значения валового продукта. Следовательно, модель можно использовать в дальнейшей работе как для экономики страны, так и для экономики региона, задавая исходные данные и получая необходимый результат – будущее значение валового продукта по каждой отрасли.

Для идентификации параметров были проведены сравнения полученных при расчетах на модели временных рядов макропоказателей экономики с соответствующими статистическими временными рядами. Таким образом, для оценки огромного числа внешних параметров нормативной модели экономики предложена новая методика их идентификации, основанная на использовании высокоскоростных параллельных вычислений.

Литература

1. Оленев Н.Н. Основы параллельного программирования в системе MPI. – М.: ВЦ РАН, 2005. – 80 с.
2. Поспелов И.Г., Поспелова И.И., Хохлов М.А., Шипулина Г.Е. Новые принципы разработки макромоделей экономики и модель современной экономики России. – М.: ВЦ РАН, 2006. – 239 с.
3. Шатров А.В. Моделирование развивающейся экономики Кировской области с помощью интеллектуальной системы ЭКОМОД/II Всероссийская научная конференция с молодежной научной школой». / Математическое моделирование развивающейся экономики ЭКОМОД-2007. Сборник трудов. – Киров: изд-во ВятГУ, 2007. – С. 259–271.
4. Оленев Н.Н., Печенкин Р.В., Чернецов А.М. Параллельное программирование в MATLAB и его приложения. – М.: ВЦ РАН, 2007. – 120 с.
5. Оленев Н.Н., Стариков А.С., Шатров А.В. Параллельные вычисления с использованием GRID – технологий в математическом моделировании региональной экономики. / Высокопроизводительные вычисления на кластерных системах. Материалы шестого Международного научно-практического семинара. Том 2. / Под ред. проф. Р.Г. Стронгина. – СПб: Изд-во СПбГУ, 2007. – С. 84–92
6. Леонтьевский сборник: жизнь, учителя, идеи. – СПб: Издательство «Нестор – История», 2006. – 384 с.

ОБ ОДНОМ АЛГОРИТМЕ ВЫЧИСЛЕНИЯ ФУНКЦИИ РАСПРЕДЕЛЕНИЯ ВЫПЛАТ В МОДЕЛИ КОЛЛЕКТИВНЫХ СТРАХОВЫХ РИСКОВ

М.В. Бацын (Нижегородский филиал государственного университета – Высшей школы экономики)

Научный руководитель – д.ф.-м.н., профессор В.А. Калягин (Нижегородский филиал государственного университета – Высшей школы экономики)

В работе исследована проблема вычисления функции распределения суммы страховых выплат при наличии эксцедентного перестрахования больших ущербов по индивидуальным страховым контрактам. Получено аналитическое выражение для этой функции при равномерном распределении ущерба в одном страховом случае. Для решения этой задачи использована однопериодная модель коллективных рисков.

Введение

Перестрахование – это специальный механизм перераспределения рисков между страховщиком и перестраховочной компанией, используемый для повышения надежности (т.е. вероятности неразорения) страховщика [1]. Для распределения рисков устанавливается так называемый уровень собственного удержания r : ущерб, не превышающий этот уровень, берет на себя сам страховщик, а все остальное оплачивает перестраховочная компания. Перестрахование индивидуальных рисков – это особый тип перестрахования, типичный для небольших страховых компаний, при котором ущерб каждого клиента (а не суммарный ущерб по всем страховым контрактам, как при стандартном перестраховании) делится между страховщиком и перестраховочной компанией согласно установленному уровню собственного удержания r [2].

Таким образом, если ущерб в i -м страховом случае описывается случайной величиной X_i с распределением $F(x)$, то случайная величина выплат страховщика Y_i имеет следующее выражение:

$$Y_i = \begin{cases} X_i, & \text{если } X_i \leq r \\ r, & \text{если } X_i > r \end{cases}$$
 . Поэтому ее распределение $F_Y(x)$

имеет разрыв в точке r . Такое распределение называется смешанным (сочетание непрерывного и дискретного) с концентрацией вероятностной массы в точке r .

При использовании перестрахования возникает задача определения оптимального уровня собственного удержания r_* , обеспечивающего максимальную надежность страховой компании, т.е. вероятность того, что суммарные выплаты по страховым случаям не превзойдут общие средства страховщика: деньги, полученные от страхователей плюс собственный резерв, но минус стоимость перестрахования [3]. Для решения этой задачи необходимо знать функцию распределения для случайной величины суммарных выплат. Однако при перестраховании индивидуальных рисков выплаты по одному страховому случаю имеют смешанное распределение, и это резко усложняет задачу получения распределения для суммы таких случайных величин. Поэтому имеет смысл использование нормального приближения для такого распределения [4]. Результаты применения нормальной аппроксимации в задаче нахождения уровня собственного удержания, максимизирующего надежность страховщика, приведены в работе [5]. Данная статья посвящена вопросу точного вычисления, и основным ее результатом является аналитическое выражение функции распределения суммарных выплат для случая равномерного распределения ущерба. Результаты, полученные с помощью нормального приближения в работе [5], сравниваются с результатами, полученными с помощью точной формулы распределения суммарных выплат.

Равномерное распределение с разрывом (со срезкой)

Рассмотрим распределение, аналогичное равномерному на отрезке $[0,1]$, но имеющее разрыв в некоторой точке r :

$$F(x) = \begin{cases} 0, & \text{если } x < 0 \\ x, & \text{если } 0 \leq x \leq r. \\ 1, & \text{если } x > r \end{cases}$$

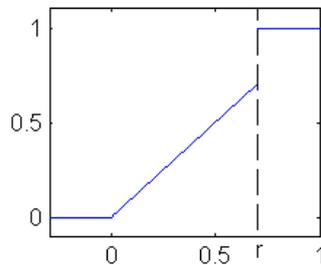


Рис. 1. Равномерное распределение со срезкой

Обозначим функцию распределения суммы из n таких случайных величин за $F_n(x)$: $S_n = X_1 + X_2 + \dots + X_n$, $X_i \sim F(x)$, $S_n \sim F_n(x)$. Будем искать общую формулу для $F_n(x)$.

Рекуррентная формула

Вероятность какого-либо события E будем обозначать как $P(E)$. Тогда по определению функции распределения: $F(x) = P(X < x)$. Сумма $n+1$ и n слагаемых X_i и их функции распределения связаны следующим образом:

$$S_{n+1} = S_n + X_{n+1}, \quad F_{n+1}(x) = P(S_n + X_{n+1} < x),$$

$$S_n \sim F_n(s) = P(S_n < s), \quad X_{n+1} \sim F(t) = P(X_{n+1} < t) \quad .$$

Из этих соотношений можно получить рекуррентную формулу $F_{n+1}(x)$ через $F_n(x)$. Функция распределения $F(x)$ имеет разрыв в точке r , поэтому сначала определимся, как это отражается на поведении $F_n(x)$.

Точка разрыва

Так как плотность распределения одной равномерно распределенной величины $f(t)$ не определена при $t = r$, так как в этой точке функция распределения имеет разрыв. X_i могут принимать значения только из отрезка $[0, r]$, так как $P(X_i < 0) = P(X_i > r) = 0$. Поэтому $S_n = X_1 + \dots + X_n \in [0, n \cdot r]$, $P(S_n < 0) = P(S_n > n \cdot r) = 0$. Значит, плотность распределения суммы $f_n(s) = 0$ при $s < 0, s > n \cdot r$. Вероятность того, что X_i равно r , будет равна величине разрыва функции распределения в точке r : $P(X_i = r) = 1 - F(r) = 1 - r$. Так как r — это максимально возможное значение X_i , то $P(S_n = n \cdot r) = P(X_1 + \dots + X_n = n \cdot r) = P(X_1 = r, X_2 = r, \dots, X_n = r) = (1 - r)^n$. Т.е. функция распределения суммы $F_n(x)$ имеет разрыв в точке $n \cdot r$ величиной $(1 - r)^n$. Таким образом, имеем:

$$P(S_n < x) = \begin{cases} 0, & \text{если } x \leq 0, x > n \cdot r \\ F_n(x), & \text{если } 0 < x < n \cdot r \\ 1 - (1 - r)^n, & \text{если } x = n \cdot r \\ 1, & \text{если } x > n \cdot r \end{cases} \quad .$$

Полученные значения будут использованы ниже при получении распределения $S_{n+1} = S_n + X_{n+1}$, как распределения суммы двух случайных величин.

Распределение суммы

Как уже было сказано выше, распределение $F_{n+1}(x)$ представляет собой следующую вероятность: $F_{n+1}(x) = P(S_n < x - X_{n+1})$. Чтобы ее найти, необходимо знать совместную плотность вероятности для величин S_n и X_{n+1} , а затем взять от нее интеграл по области $D: S_n < x - X_{n+1}$. Случайные величины величин S_n и X_{n+1} являются независимыми. Поэтому их совместная плотность распределения равна произведению плотностей $f_n(s)$ и $f(t)$. Однако, функции распределения имеют разрывы в точках $n \cdot r$ и r соответственно, поэтому значения $s = n \cdot r$ и $t = r$ необходимо рассматривать отдельно. Рассмотрим область значений переменных s и t на плоскости (рис. 2).

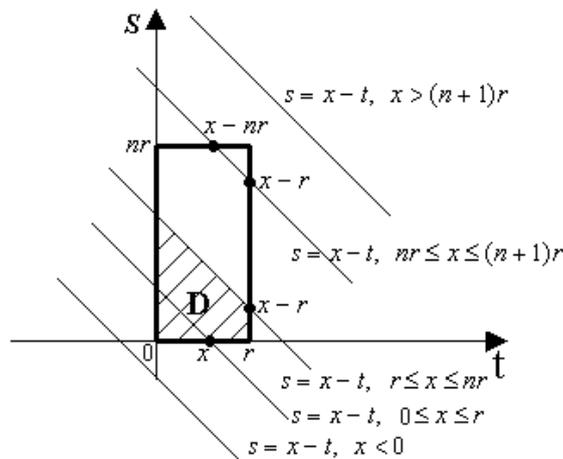


Рис. 2. Область значений S_n и X_{n+1}

На рис. 2 границы области допустимых значений s и t выделены жирными линиями. Искомая вероятность $F_{n+1}(x) = P(S_n < x - X_{n+1})$ равна интегралу от совместной плотности распределения по области, находящейся под прямой $s = x - t$. Поэтому из рисунка видно, что: $F_{n+1}(x) = 0$, если $x < 0$, $F_{n+1}(x) = 1$, если $x > (n+1)r$. Распишем $F_{n+1}(x)$ в общем случае, учитывая особые значения $s = n \cdot r$ и $t = r$:

$$P(S_n < x - X_{n+1}) = \iint_D f_n(s) \cdot f(t) ds dt + P(X_{n+1} = r) \cdot P(S_n < x - r, S_n \neq n \cdot r) + \\ + P(S_n = n \cdot r) \cdot P(X_{n+1} < x - n \cdot r, X_{n+1} \neq r) + P(S_n = n \cdot r) \cdot P(X_{n+1} = r) \cdot P(x > (n+1)r)$$

В последнем выражении:

$$P(x > (n+1)r) = \begin{cases} 1, & \text{если } x > (n+1)r \\ 0, & \text{иначе} \end{cases}$$

Теперь все вероятности расписаны так, что вместо них можно подставлять значения функций распределения:

$$F_{n+1}(x) = \iint_D f_n(s) \cdot f(t) ds dt + (1-r) \cdot F_n(x-r) + (1-r)^n \cdot F(x-n \cdot r) + \\ + (1-r)^n \cdot (1-r) \cdot P(x > (n+1)r)$$

Заметим, что $F_{n+1}(x) = 1$, если $x > (n+1)r$. Кроме того, если $x < n \cdot r$, то $F(x - n \cdot r) = 0$. Поэтому при $nr \leq x \leq (n+1)r$

$$F_{n+1}(x) = \iint_D f_n(s) \cdot f(t) ds dt + (1-r) \cdot F_n(x-r) + (1-r)^n \cdot (x-nr),$$

а при $0 \leq x \leq nr$

$$F_{n+1}(x) = \iint_D f_n(s) \cdot f(t) ds dt + (1-r) \cdot F_n(x-r).$$

Остается найти интеграл по области D . Рассмотрим сначала случай $0 \leq x \leq nr$, который делится на два случая: первый – $r \leq x \leq nr$ (см. рис. 2):

$$\iint_D f_n(s) \cdot f(t) ds dt = \int_0^r dt \int_0^{x-t} f_n(s) ds = \int_0^r F_n(x-t) dt = \int_{x-r}^x F_n(t) dt.$$

и второй – $0 \leq x \leq r$:

$$\iint_D f_n(s) \cdot f(t) ds dt = \int_0^x dt \int_0^{x-t} f_n(s) ds = \int_0^x F_n(x-t) dt = \int_0^x F_n(t) dt.$$

Заметим также, что в этом случае $x-r < 0$, и поэтому $F_n(x-r) = 0$, так что выражение для $F_{n+1}(x)$ еще сильнее упрощается. В случае $nr \leq x \leq (n+1)r$ интеграл будет представлять собой сумму 2 интегралов, т.к. область D разбивается на 2 (см. пунктирную линию на рис. 3):

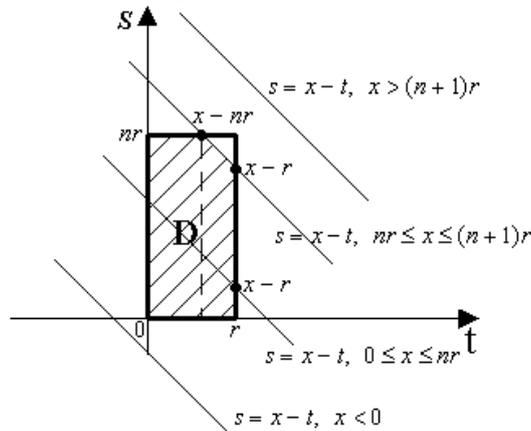


Рис. 3. Область значений S_n и X_{n+1}

$$\iint_D f_n(s) \cdot f(t) ds dt = \int_0^{x-nr} dt \int_0^{nr} f_n(s) ds + \int_{x-nr}^r dt \int_0^{x-t} f_n(s) ds = (1-(1-r)^n) \cdot (x-nr) + \int_{x-r}^{nr} F_n(t) dt.$$

В результате получаем рекуррентную формулу для функции распределения суммы:

$$F_{n+1}(x) = \begin{cases} 0, & \text{если } x < 0 \\ \int_0^x F_n(t) dt, & \text{если } 0 \leq x \leq r \\ \int_{x-r}^x F_n(t) dt + (1-r) \cdot F_n(x-r), & \text{если } r \leq x \leq nr \\ \int_{x-r}^{nr} F_n(t) dt + (x-nr) + (1-r) \cdot F_n(x-r), & \text{если } nr \leq x \leq (n+1)r \\ 1, & \text{если } x > (n+1)r \end{cases}.$$

Из этой формулы видно, что функция состоит из нескольких кусков, так как на разных отрезках имеет разные выражения.

Распределение суммы – кусочная функция

Нетрудно доказать, что $F_{n+1}(x)$ на $[0, (n+1)r]$ состоит из $n+1$ части ($k=2, n$):

$$F_{n+1}(x) = \begin{cases} F_{n+1}^1(x) = \int_0^x F_n^1(t) dt, & \text{если } 0 \leq x \leq r \\ F_{n+1}^k(x) = \int_{x-r}^{(k-1)r} F_n^{k-1}(t) dt + \int_{(k-1)r}^x F_n^k(t) dt + (1-r) \cdot F_n^{k-1}(x-r), & \text{если } (k-1)r \leq x \leq kr \\ F_{n+1}^{n+1}(x) = \int_{x-r}^{nr} F_n^n(t) dt + (x-nr) + (1-r) \cdot F_n^n(x-r), & \text{если } nr \leq x \leq (n+1)r \end{cases}$$

Данная формула будет использована далее для получения общей формулы $F_n(x)$, но сначала выведем вспомогательную формулу для $F_n^n(x)$.

Вспомогательная формула

Справедлива следующая формула на отрезке $[(n-1)r, nr]$ для функции распределения суммы n величин, распределенных равномерно со срезкой:

$$F_n^n(x) = 1 - (-1)^n \cdot \sum_{i=0}^n C_n^i (r-1)^i \frac{(x-nr)^{n-i}}{(n-i)!}. \quad (1)$$

Эту формулу нетрудно доказать по методу математической индукции, воспользовавшись рекуррентной зависимостью, найденной ранее.

Общая формула

Для k -й части (для $x \in [(k-1)r, kr]$) функции распределения суммы n случайных величин, имеющих равномерное распределение на $[0, 1]$ со срезкой в точке r , справедлива следующая общая формула:

$$F_n^k(x) = \sum_{i=0}^{k-1} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j}}{(n-j)!} \right].$$

Для сокращения формул введем следующие обозначения:

$$V_{i,j}(x) = (r-1)^j \frac{(x-ir)^{n-j+1}}{(n-j+1)!}, \quad U_i(x) = (-1)^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j+1}}{(n-j+1)!}.$$

Доказательство для n -й части функции распределения

Сначала докажем формулу для $k = n$, т.е. для $F_n^n(x)$ (это наиболее сложный случай), снова используя метод математической индукции. Справедливость для $n = 1$ очевидна. Считая формулу верной для F_n^n , докажем ее справедливость для F_{n+1}^{n+1} . Итак, известно:

$$F_n^n(x) = \sum_{i=0}^{n-1} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j}}{(n-j)!} \right].$$

Приравняв правые части этого выражения и выражения (1), приходим к равенству:

$$1 = \sum_{i=0}^n \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j}}{(n-j)!} \right]. \quad (2)$$

Для доказательства индукционного перехода используем рекуррентное соотношение.

Представим $(x-nr)$ в виде $\int_{nr}^x 1 dt$ и заменим единицу на выражение (2):

$$(x - nr) = \sum_{i=0}^n C_n^i U_i(x) - \sum_{i=0}^n C_n^i U_i(nr).$$

Пользуясь предположением индукции для F_n^n и вычисляя интеграл, запишем 1-е слагаемое рекуррентной формулы:

$$\int_{x-r}^{nr} F_n^n(t) dt = \sum_{i=0}^{n-1} C_n^i U_i(nr) - \sum_{i=0}^{n-1} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j V_{i+1,j}(x) \right].$$

И, наконец, последнее слагаемое:

$$(1-r) \cdot F_n^n(x-r) = - \sum_{i=0}^{n-1} \left[(-1)^i C_n^i \sum_{j=1}^{i+1} C_i^{j-1} V_{i+1,j}(x) \right].$$

В результате

$$F_{n+1}^{n+1}(x) = \sum_{i=0}^n C_n^i U_i(x) - \sum_{i=0}^n C_n^i U_i(nr) + \sum_{i=0}^{n-1} C_n^i U_i(nr) - \\ - \sum_{i=0}^{n-1} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j V_{i+1,j}(x) \right] - \sum_{i=0}^{n-1} \left[(-1)^i C_n^i \sum_{j=1}^{i+1} C_i^{j-1} V_{i+1,j}(x) \right].$$

Заметим, что $U_n(nr) = 0$ и, следовательно, $\sum_{i=0}^n C_n^i U_i(nr) = \sum_{i=0}^{n-1} C_n^i U_i(nr)$. После преобразований над суммами получаем:

$$F_{n+1}^{n+1}(x) = C_{n+1}^0 U_0(x) + \sum_{i=1}^n C_{n+1}^i U_i(x) = \sum_{i=0}^n C_{n+1}^i U_i(x) = \sum_{i=0}^n (-1)^i C_{n+1}^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j+1}}{(n-j+1)!}$$

Получили требуемое выражение для F_{n+1}^{n+1} , а значит, по индукции формула справедлива для любого n . Итак, общая формула доказана для частного случая $k = n$. Рассмотрим еще один частный случай – $k = 1$, так как рекуррентное соотношение имеет особый вид не только для $F_{n+1}^{n+1}(x)$, но и для $F_{n+1}^1(x)$.

Доказательство для первой части функции распределения

Случай $k = 1$ – самый простой, и формула, которую необходимо доказать, достаточно проста: $F_n^1 = x^n / n!$. Для $n = 1$ эта формула выполняется. Считая формулу верной для F_n^1 , покажем, что она справедлива и для F_{n+1}^1 , используя рекуррентное соотношение:

$$F_{n+1}^1 = \int_0^x F_n^1(t) dt = \int_0^x \frac{t^n}{n!} dt = \frac{t^{n+1}}{(n+1)!} \Big|_0^x = \frac{x^{n+1}}{(n+1)!}.$$

Таким образом, база индукции и индуктивный переход выполняются, а значит, формула доказана. Теперь остается только доказать общий случай $k = 2, 3, \dots, n-1$.

Общий случай

Для $k = 2, 3, \dots, n-1$ будем доказывать общую формулу индукцией по n . Заметим, что $F_n^k(x)$ имеет ровно n частей на отрезке $[0, nr]$, и значит, в общей формуле $n \geq k$. Тогда база индукции – $n = 2$, а значит, $k = 1, 2$. Формула для F_2^1 подпадает под доказательство в параграфе «Доказательство для первой части функции распределения», а формула для F_2^2 – под доказательство в параграфе «Доказательство для n -й части функции распределения». Теперь, считая, что формула верна для всех F_n^k ($n \geq 2$) при

$k = 1, 2, 3, \dots, n$, докажем ее для всех F_{n+1}^k при $k = 2, 3, \dots, n-1$ (случаи $k = 1$ и $k = n+1$ уже были доказаны выше). Итак, имеем:

$$F_n^k(x) = \sum_{i=0}^{k-1} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j}}{(n-j)!} \right], \quad F_n^{k-1}(x) = \sum_{i=0}^{k-2} \left[(-1)^i C_n^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n-j}}{(n-j)!} \right].$$

Для доказательства формулы F_{n+1}^k используем рекуррентное соотношение:

$$F_{n+1}^k(x) = \int_{x-r}^{(k-1)r} F_n^{k-1}(t) dt + \int_{(k-1)r}^x F_n^k(t) dt + (1-r) \cdot F_n^{k-1}(x-r).$$

Выпишем каждое из слагаемых, используя формулы для F_n^k и F_n^{k-1} , верные по предположению индукции:

$$\begin{aligned} \int_{x-r}^{(k-1)r} F_n^{k-1}(t) dt &= \sum_{i=0}^{k-2} C_n^i U_i((k-1)r) - \sum_{i=0}^{k-2} (-1)^i C_n^i \sum_{j=0}^i C_i^j V_{i+1,j}(x), \\ \int_{(k-1)r}^x F_n^k(t) dt &= \sum_{i=0}^{k-1} C_n^i U_i(x) - \sum_{i=0}^{k-1} C_n^i U_i((k-1)r), \quad (1-r)F_n^{k-1}(x-r) = - \sum_{i=0}^{k-2} (-1)^i C_n^i \sum_{j=1}^{i+1} C_i^{j-1} V_{i+1,j}(x). \end{aligned}$$

Замечая, что $U_{k-1}((k-1)r) = 0$, а следовательно, $\sum_{i=0}^{k-1} C_n^i U_i((k-1)r) = \sum_{i=0}^{k-2} C_n^i U_i((k-1)r)$, получаем после сложения и преобразований:

$$F_{n+1}^k(x) = \sum_{i=1}^{k-1} C_n^i U_i(x) + \sum_{i=1}^{k-1} C_n^{i-1} U_i(x) + U_0(x) = \sum_{i=0}^{k-1} \left[(-1)^i C_{n+1}^i \sum_{j=0}^i C_i^j (r-1)^j \frac{(x-ir)^{n+1-j}}{(n+1-j)!} \right].$$

Получив требуемую формулу, мы доказали индукционный переход, а значит, и всю формулу для всех n и k .

Точная функция распределения в модели коллективных рисков

Согласно модели коллективных рисков на коротком интервале времени, число страховых случаев Q имеет пуассоновское распределение с параметром λ (среднее число страховых случаев за год), и вероятность $Q = n$ будет $P(Q = n) = e^{-\lambda} \cdot \lambda^n / n!$. Функция распределения суммарных выплат по всем страховым случаям, или вероятность того, что выплаты не превзойдут значение x , складывается из функций распределения для каждого значения числа страховых случаев Q от нуля до бесконечности, умноженных на соответствующие вероятности $P(Q = n)$:

$$G(x) = P(Y < x) = \sum_{n=0}^{\infty} P(Q = n) \cdot F_n(x) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \cdot F_n(x)$$

Эта сумма разбивается на две, так как $F_n(x) = 1$ при $x > n \cdot r$, т.е. $n \leq [x/r]$:

$$G(x) = \sum_{n=0}^{[x/r]} \frac{\lambda^n}{n!} e^{-\lambda} + \sum_{n=[x/r]+1}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \cdot F_n^k(x), \quad k = [x/r] + 1$$

Если оборвать эти суммы на m -м слагаемом, то остатки R_m можно оценить следующим значением:

$$R_m \leq e^{-\lambda} \cdot \frac{\lambda^{m+1}}{(m+1)!} \cdot \left[1 + \frac{\lambda}{m+2} + \frac{\lambda^2}{(m+2)(m+3)} + \dots \right] < e^{-\lambda} \cdot \frac{\lambda^{m+1}}{(m+1)!} \cdot \left[1 + \frac{\lambda}{m} + \left(\frac{\lambda}{m} \right)^2 + \dots \right]$$

Наложим условие сходимости на знаменатель этой геометрической прогрессии: $m > \lambda$. Тогда, пользуясь формулой суммы бесконечного геометрического ряда, получим:

$$R_m < e^{-\lambda} \cdot \frac{\lambda^{m+1}}{(m+1)!} \cdot \frac{1}{1-\lambda/m} = e^{-\lambda} \cdot \frac{\lambda^{m+1}}{(m+1)!} \cdot \frac{m}{m-\lambda}$$

Потребовав выполнение следующих неравенств на $m : m > \lambda$, $e^{-\lambda} \cdot \frac{\lambda^{m+1}}{(m+1)!} \cdot \frac{m}{m-\lambda} < \varepsilon$,

мы найдем количество слагаемых m , которое достаточно, чтобы получить значение суммы с точностью ε . Например:

$$\lambda = 1 : m = 10 \Rightarrow \varepsilon = 10^{-8}, \quad \lambda = 10 : m = 20 \Rightarrow \varepsilon = 0.001, \quad m = 40 \Rightarrow \varepsilon = 10^{-13},$$

$$\lambda = 20 : m = 30 \Rightarrow \varepsilon = 0.01, \quad m = 50 \Rightarrow \varepsilon = 10^{-9}.$$

Таким образом, для вычисления данной бесконечной суммы достаточно посчитать лишь небольшое число слагаемых.

Надежность страховой компании

Надежность страховой компании – это вероятность того, что суммарные выплаты во всех страховых случаях Y не превзойдут сумму средств, полученных страховщиком от клиентов в качестве страховых взносов $S = N \cdot \Pi \cdot (1 + \theta)$ (N – число страховых контрактов, Π – рисковая премия, θ – рисковая надбавка) минус плата за перестрахование $\tilde{S} = N \cdot \tilde{\Pi} \cdot (1 + \alpha + \beta)$ ($\tilde{\Pi}$ – рисковая премия, α – рисковая надбавка, β – нагрузка на ведение дела, устанавливаемые перестраховочной компанией). Таким образом, надежность равна: $REL = P(Y < N \cdot \Pi \cdot (1 + \theta) - N \cdot \tilde{\Pi} \cdot (1 + \alpha + \beta)) = P(Y < S - \tilde{S})$. Т.е. надежность равна значению функции распределения выплат в точке $S - \tilde{S} : G(S - \tilde{S})$. Заметим, что рисковая премия $\tilde{\Pi}$ зависит от r , и слишком малое значение r (т.е. риска, оставляемого страховщиком на собственном удержании) дает слишком большое значение $\tilde{\Pi}$, и у страховщика может просто не хватить средств на оплату такого перестрахования. Сама функция распределения $G(x)$ также зависит от r , и для любого его значения можно посчитать надежность REL . Таким образом, получаем функцию $REL(r)$. Точка r^* , которой соответствует наибольшее значение надежности $REL(r^*)$ будет оптимальным уровнем собственного удержания. Надежность, полученная с помощью нормальной аппроксимации, сравнивается с точной функцией $REL(r)$ в разделе «Результаты».

Вычислительная сложность

Рассматривая формулу для функции распределения суммарных выплат страховой компании, можно понять, что ее вычисление на компьютере будет занимать значительное время. Это время быстро растет с увеличением числа точек x , числа различных значений параметра r , с повышением точности ε , для которых требуется получить значение функции распределения.

Для решения этой проблемы целесообразно распараллелить алгоритм вычисления и использовать многопроцессорные компьютеры или кластеры из нескольких компьютеров. Здесь возможны 3 различных случая, возникающие в страховании.

1. Перестраховочная компания сама определяет уровень собственного удержания r , и для страховщика этот параметр фактически является константой. Но страховщик может влиять на объем своих средств, полученных по страховым контрактам, например, привлекая клиентов различными способами. Значит, его интересует функция распределения выплат при различных значениях x . Таким образом, значение r фиксировано, точность ε в общем-то тоже фиксирована, и функцию распределения нужно вычислять только в различных точках x .

2. Перестраховочная компания не фиксирует уровень собственного удержания r , а лишь определяет стоимость перестрахования в зависимости от r . Страховщик уже обладает определенным набором клиентов и знает объем своих средств x . Он может выбирать различные значения r , например для оптимизации своей надежности. Тогда функцию распределения суммарных страховых выплат необходимо вычислить для множества значений r .
3. И объем средств x , полученных от клиентов, и уровень r фиксированы, но среднее число страховых случаев λ велико. Страховой компании необходимо знать с большой точностью ε риск разорения, т.е. значение функции распределения. Это нужно, например, чтобы определить объем средств, которые должны находиться в страховом резерве. Таким образом, вычислить функцию надо лишь в одной точке, но с высокой точностью и при большом числе слагаемых в сумме, возникающем за счет большого λ .

Каждый из этих случаев предоставляет разные возможности для организации параллельных вычислений.

Равномерное распараллеливание

В 1-м и 2-м описанных случаях организовать параллельные вычисления получается наиболее просто и оптимально. Дело в том, что значение функции при одних значениях параметров x и r никак не зависит от ее значения при других x и r . Поэтому нетрудно распараллелить алгоритм вычисления по этим параметрам. Т.е., если необходимо вычислить функцию для N вариантов значений x и r , а число процессоров или компьютеров в кластере – M , то нужно создать M потоков и отдать каждому вычисление функции в N/M точках. При этом распараллеливание будет равномерным, т.е. все потоки закончат вычисления примерно в одно время. Равномерность обеспечивается тем, что вычисление функции распределения суммы ущерба F при одних значениях параметров x и r ничем не отличается от вычисления при других x и r .

Если число различных значений x и r велико, то даже многопоточный алгоритм будет работать значительное время. Вычисления можно оптимизировать, если заметить рекуррентную зависимость между величинами $F_n^{k+1}(x)$ и $F_n^k(x)$:

$$F_n^{k+1}(x) = F_n^k(x) + (-1)^k C_n^k \sum_{j=0}^k C_k^j (r-1)^j \frac{(x-kr)^{n-j}}{(n-j)!} = F_n^k(x) + (-1)^k C_n^k \sum_{j=0}^k a_j = F_n^k(x) + \Delta_n^k(x)$$

Т.е. для каждого значения n достаточно сначала вычислить $F_n^1(x) = x^n/n!$, а остальные $F_n^{k+1}(x)$ вычислять простым сложением с $\Delta_n^k(x)$. Кроме того, слагаемые a_j можно вычислять эффективно, если заметить, что:

$$a_0 = (x-kr)^n/n!, \quad a_{j+1} = a_j \cdot (k-j)(n-j)/(j+1) \cdot (r-1)/(x-kr)$$

Причем множитель $(r-1)/(x-kr)$ нужно вычислить только один раз, так как он не зависит от j . Используя такие оптимизации, удастся сократить общее время вычислений.

Распараллеливание по параметру n

Приведенное выше равномерное распараллеливание невозможно в 3-м случае, поэтому необходимо организовывать разделение вычислений по параметру n . Вследствие оптимизаций появляется зависимость между величинами $F_n^{k+1}(x)$ и $F_n^k(x)$, и поэтому распараллеливание по k невозможно. В результате получаются следующие независимые цепочки вычислений:

$$F_1^1, F_2^1 \rightarrow F_2^2, F_3^1 \rightarrow F_3^2 \rightarrow F_3^3, \dots, F_m^1 \rightarrow F_m^2 \rightarrow \dots \rightarrow F_m^m$$

Все цепочки различаются по длине, а следовательно, и по времени вычисления. Здесь необходимо использовать стратегию пула потоков, чтобы максимально уравнивать вычисления каждого из потоков друг с другом. Чем более равномерно будут загружены потоки, тем меньше будет общее время вычислений, потому что оно равно времени работы самого загруженного потока. Ясно, что потоков в пуле должно равняться числу процессоров или компьютеров в кластере. Каждый поток должен брать самую длинную цепочку из оставшихся для вычисления до тех пор, пока не останется ни одной не вычисленной цепочки.

Оптимизированное распараллеливание по параметру n

Влияния зависимости между величинами $F_n^{k+1}(x)$ и $F_n^k(x)$ на вычисления можно избежать, если заставить потоки из пула вычислять не $F_n^k(x)$, а лишь $\Delta_n^k(x)$, между которыми нет никаких зависимостей. Поскольку вычисление $\Delta_n^k(x)$ тем дольше, чем больше значение k , то каждый поток должен вычислять $\Delta_n^k(x)$ с наибольшим значением k из оставшихся до тех пор, пока не останется ни одной не вычисленной $\Delta_n^k(x)$. Первый поток должен сначала вычислить $F_n^1(x)$ для всех значений n , а затем, когда будут вычислены все $\Delta_n^k(x)$, должен получить значения всех $F_n^k(x)$ простым сложением с $\Delta_n^k(x)$. Такой алгоритм дает наиболее равномерное распределение нагрузки между потоками и наименьшее общее время исполнения.

Результаты

Для получения оптимального с точки зрения надежности страховой компании уровня собственного удержания r^* строится функция зависимости надежности от r : $REL(r)$. Применяя нормальную аппроксимацию, мы получаем функцию надежности неточно. Обозначим эту неточную функцию $REL_N(r)$. Из нее определяется оптимальное значение уровня собственного удержания r_N^* . В данной работе найдена точная функция надежности $REL(r)$, из которой можно получить точное значение r^* . Сравнивая значения точной функции $REL(r)$ в точках r^* и r_N^* , мы видим, насколько страховая компания проигрывает (ошибается) в надежности: $REL(r^*) - REL(r_N^*)$, используя нормальное приближение. Ниже приведены результаты для различных значений параметра λ в модели коллективных рисков.

Результаты для $\lambda = 1$

На графике приведены зависимости надежности страховщика от выбранного уровня собственного удержания. Синим цветом (график без скачков) изображена функция, полученная с использованием нормального приближения. Красным цветом показана точная функция надежности, полученная с помощью аналитической формулы для распределения случайной величины суммы выплат.

Оптимальное значение уровня собственного удержания, найденное с помощью нормальной аппроксимации – $r_N^* = 0.31$, точной функции распределения – $r^* = 0.37$. Проигрыш в надежности $REL(r^*) - REL(r_N^*) = 0.93205 - 0.93149 = 0.00056$.

Таким образом, применяя нормальное приближение (более удобное в расчетах и подходящее для любых распределений ущерба в страховом случае) в данном примере, страховая компания совершит ошибку в надежности $5 \cdot 10^{-4}$, что вполне допустимо. Заметим, что в этом примере точная зависимость надежности достаточно сильно отлича-

ется от неточной и имеет сильный скачок. Если бы максимум неточной функции надежности оказался правее этого скачка, то страховая компания совершила бы серьезную ошибку 4–5%. Однако с ростом среднего числа страховых случаев λ эта разница снижается, а разрывы сглаживаются или вообще исчезают. Поэтому нормальное приближение разумно использовать при больших значениях λ , как в следующих примерах

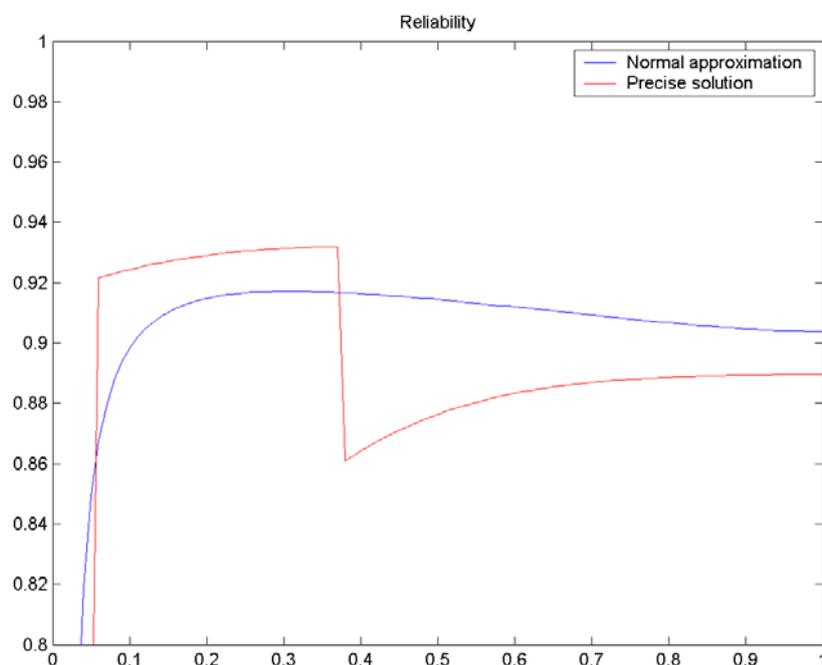


Рис. 4. Надежность при $\lambda = 1$

Результаты для $\lambda = 30$

На рис. 5 хорошо видно, что точность нормальной аппроксимации улучшается с ростом λ . График зависимости надежности от r , полученный с ее помощью, повторяет форму точной функции $REL(r)$.

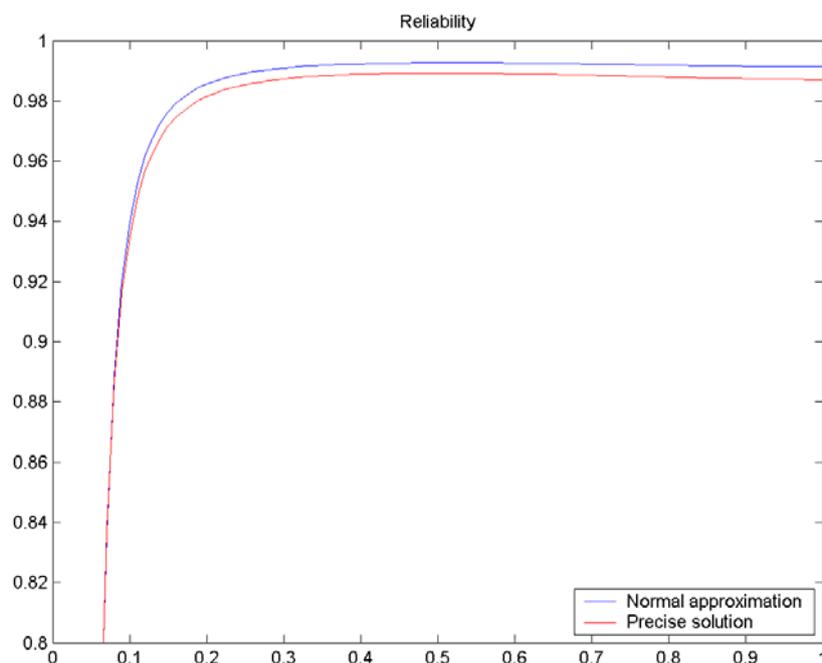


Рис. 5. Надежность при $\lambda = 30$

Оптимальные значения: $r_N^* = 0.58$, $r^* = 0.56$. Проигрыш в надежности: $REL(r^*) - REL(r_N^*) = 0.96837 - 0.96834 = 0.00003$.

Заключение

В теории вероятности существует множество различных распределений случайных величин, простых и сложных. Но когда возникает необходимость получить распределение суммы одинаково распределенных величин, задача резко усложняется. Формула распределения суммы найдена в теории вероятности лишь для нескольких несложных распределений случайной величины: равномерного, нормального, пуассоновского и других. Поэтому вывод такой формулы для какого-либо распределения можно считать большой удачей.

В данной работе получено аналитическое выражение для распределения суммы случайных величин, имеющих равномерное распределение со срезкой. Сложность такого распределения заключается в том, что оно имеет разрыв в некоторой точке r . Это распределение возникает в актуарной математике в задачах, связанных с экседентным перестрахованием индивидуальных рисков. В работе рассмотрено его применение в проблеме оптимизации надежности страховой компании.

В актуарной математике известен лишь один способ точного вычисления функции распределения суммы страховых выплат при наличии перестрахования. Это рекуррентная формула Пейнджера [2, 6]. Но область ее применения ограничена 2 условиями.

1. Распределение числа страховых случаев должно удовлетворять особому рекуррентному условию, под которое подходят 3 общеизвестных распределения: распределение Пуассона, биномиальное и отрицательное биномиальное распределения.
2. Распределение ущерба в одном страховом случае должно быть дискретным. Это условие заставляет аппроксимировать непрерывное распределение дискретным, теряя в точности решения. При большой частоте дискретизации вычисления занимают значительное время.

Литература

1. Бауэрс Н., Гербер Х., Хикман Дж., Джонс Д., Несбитт С. Актуарная математика – М.: Вильямс, 2001. – 1020 с.
2. Kremer E. Applied risk theory – Shaker Verlag, 1999. – 240 с.
3. Голубин А.Ю. Математические модели в теории страхования: построение и оптимизация – М.: Анкил, 2003. – 160 с.
4. Корнилов И.А. Основы страховой математики – М.: Юнити-Дана, 2004. – 400 с.
5. Бацын М.В., Калягин В.А. Определение оптимального уровня собственного удержания при экседентном перестраховании убытка // Известия АИН им. А.М. Прохорова, серия Бизнес-информатика. – 2005. – № 12. – С. 67-74.
6. Panjer H.H., Willmot G.E. Insurance risk models – Schaumburg/Illinois: Society of Actuaries, 1992. – 360 с.

МУЛЬТИАГЕНТНАЯ СИСТЕМА ДИНАМИЧЕСКОЙ БАЛАНСИРОВКИ РАСПРЕДЕЛЕННОЙ ИМИТАЦИОННОЙ МОДЕЛИ

А.А. Козлов (Пермский государственный университет)
Научный руководитель – к.ф.-м.н., доцент Е.Б. Замятина
(Пермский государственный университет)

При выполнении сложных распределенных имитационных моделей часто приходится сталкиваться с неравномерностью нагрузки на вычислительные узлы, что может привести к большим временным затратам при выполнении имитационного эксперимента. В работе предложен метод распределения нагрузки, использующий интеллектуальных агентов.

Введение

В настоящее время существует тенденция широкого использования высокопроизводительных вычислений, соответственно возникает необходимость в применении программных систем, поддерживающих распределенное имитационное моделирование. Распределенное имитационное моделирование предполагает, что компоненты имитационной модели располагаются на различных вычислительных узлах, используя при выполнении имитационного эксперимента вычислительные ресурсы сразу нескольких узлов компьютерной сети (или многопроцессорной ЭВМ). Компоненты имитационной модели обмениваются сообщениями, которые передаются по линиям связи между вычислительными узлами.

Однако гетерогенность вычислительных узлов и линий связи между ними, а также гетерогенность самой имитационной модели могут привести к возникновению дисбаланса нагрузки. При возникновении дисбаланса часть вычислительных узлов может простаивать, в то время как другие будут перегружены, и компоненты имитационной модели будут ожидать освобождения вычислительных ресурсов в течение непредсказуемого интервала времени. Следовательно, выигрыш от применения вычислительных ресурсов всех узлов вычислительной системы, на которой проводится имитационный эксперимент, сводится к нулю. В связи с этим возникает необходимость в динамическом перераспределении ресурсов, перемещении компонентов модели с одного узла на другие с целью восстановления баланса. Восстановление баланса позволит сократить временные затраты и в конечном итоге приведет к сокращению времени проведения имитационного эксперимента.

Разработано большое количество алгоритмов и стратегий балансировки (например, RCL стратегия). Однако методы восстановления баланса, предложенные различными авторами, работают оптимально только для конкретных имитационных моделей. Попытки создать универсальный алгоритм лишь частично решают проблему. Суть в том, что все эти алгоритмы не учитывают специфические особенности имитационной модели.

На кафедре математического обеспечения ВС Пермского государственного университета ведутся работы по созданию распределенной системы имитации. Подсистема балансировки является частью этой системы. Чтобы избежать недостатков, присущих ранее рассмотренным подсистемам балансировки, предложено реализовать балансировку, основанную на знаниях. Для различных классов имитационных моделей разработаны правила, учитывающие особенности имитационных моделей этого класса, топологию имитационной модели и компьютерной сети, на которой выполняется имитационный эксперимент, и т.д. Правила хранятся в базе знаний и управляют алгоритмом балансировки [2].

Однако управление является централизованным, все правила хранятся в единой базе знаний.

Чтобы ресурсы распределенной вычислительной среды в наибольшей степени были использованы с целью оптимизации времени выполнения алгоритма балансировки, было предложено использовать мультиагентный подход. Агенты успешно используются в распределенных системах, обеспечивают масштабирование и расширение функциональности системы и самоорганизуются [3].

Рассмотрим более подробно представление имитационной модели, постановку задачи балансировки и принятые решения.

Представление имитационной модели

Описание имитационной модели в Triad состоит из трех слоев: слоя структур (STR), слоя рутин (ROUT) и слоя сообщений (MES). Таким образом, модель в системе Triad можно определить как $M = \{STR, ROUT, MES\}$.

Слой структур представляет собой совокупность объектов, взаимодействующих друг с другом посредством посылки сообщений. Каждый объект имеет полюсы (входные Pin и выходные Pout), которые служат соответственно для приема и передачи сообщений. Слой структур можно представить графом. В качестве вершин графа следует рассматривать отдельные объекты. Дуги графа определяют связи между объектами.

Объекты действуют по определенному алгоритму поведения, который описывают с помощью рутины (rout). Рутинa представляет собой последовательность событий e_i , планирующих друг друга ($e_i \in E, i=1 \div n, E$ – множество событий, множество событий рутины является частично упорядоченным в модельном времени). Выполнение события сопровождается изменением состояния q_k объекта. Состояние объекта определяется значениями переменных var_j рутины ($var_j \in Var, j=1 \div m, Var$ – множество переменных рутины). Таким образом, система имитации является событийно-ориентированной. Рутинa, также как и объект, имеет входные (Prin и выходные Prout) полюса. Входные полюса служат, соответственно, для приема сообщений, выходные полюса – для их передачи. В множестве событий рутины выделено входное событие e_{in} . Все входные полюса рутины обрабатываются входным событием. Обработка выходных полюсов осуществляется остальными событиями рутины. Для передачи сообщения служит специальный оператор out (`out <сообщение> through <имя полюса>`). Совокупность рутин определяет слой рутин ROUT.

Слой сообщений (MES) предназначен для описания сообщений сложной структуры. Система Triad реализована таким образом, что пользователь может описать только один слой. Так, если возникает необходимость в исследовании структурных особенностей модели, то можно описать в модели только слой структур.

Алгоритмом имитации называют объекты, функционирующие по определенным сценариям, и синхронизирующий их алгоритм.

Для сбора, обработки и анализа имитационных моделей в системе Triad.Net существуют специальные объекты – информационные процедуры и условия моделирования. Информационные процедуры и условия моделирования реализуют алгоритм исследования.

Информационные процедуры ведут наблюдение только за теми элементами модели (событиями, переменными, входными и выходными полюсами), которые указаны пользователем. Если в какой-нибудь момент времени имитационного эксперимента пользователь решит, что следует установить наблюдение за другими элементами или выполнять иную обработку собираемой информации, он может сделать соответствующие указания, подключив к модели другой набор информационных процедур. Информационные процедуры являются единственным средством системы для одновременно-

го доступа к элементам модели, принадлежащим разным объектам. Именно с помощью информационных процедур пользователь может осуществить взаимодействие с объектами модели во время имитации.

Условия моделирования анализируют результат работы информационных процедур и определяют, выполнены ли условия завершения моделирования.

Подсистема анализа модели обеспечивает получение информации по заранее сформулированному запросу, а не ограничивает пользователя строго регламентированным набором собираемых данных. Такой подход к сбору информации позволяет избежать избыточности собранной информации или того, что она окажется недостаточной.

Информационные процедуры и условия моделирования используют и для сбора информации о поведении модели, о ее характеристиках и в подсистеме балансировки для сбора информации о модели.

Задача балансировки

Задача балансировки ставится как задача отображения неизоморфных связных графов, $V: TM \rightarrow NG$, где TM – множество графов моделей, NG – множество графов – конфигураций компьютерной сети. Граф $G \in NG$, $G = \{C, Ed\}$, определяется множеством вычислительных узлов C и множеством ребер Ed , обозначающих линии связи, NG как граф, содержащий все возможные (допустимые) графы G в качестве подграфов. Граф $M \in TM$, $M = \{U, V, W\}$, задает имитационную модель, U – множество входных и выходных полюсов, V – множество вершин, которые отображают отдельные компоненты имитационной модели, W – связи между этими компонентами. Следует отметить, что одному вычислительному узлу $C_j \in G$ может соответствовать любое число вершин (в том числе и нуль) вершин графа M .

Итак, необходимо расположить объекты имитационной модели по узлам вычислительной сети таким образом, чтобы нагрузка на вычислительных узлах была приблизительно равной, при этом затраты на коммуникацию между узлами также должна быть минимизирована. Перераспределение ресурсов происходит динамически, т.е. в процессе выполнения имитационного прогона.

Алгоритм балансировки должен быть оптимален для любой имитационной модели. Однако создание такого алгоритма является достаточно сложной задачей, так как данный алгоритм должен учитывать специфику работы каждой конкретной модели. Имитационные модели могут значительно отличаться друг от друга своими характеристиками: структурой, интенсивностью обмена информацией между отдельными компонентами, алгоритмами поведения этих компонентов. В работе предложен метод балансировки, основанный на интеллектуальных агентах. Система балансировки является подсистемой системы имитации Triad.Net и состоит из множества интеллектуальных агентов нескольких типов (на каждом вычислительном узле находятся по одному агенту каждого типа), каждый из которых решает свою собственную задачу балансировки нагрузки узла на котором он сам находится, а вместе они решают общую задачу балансировки.

Данный подход основан на расширяемых знаниях агентов о сети и модели имитации, что позволяет выполнять балансировку для различных моделей имитации и различных топологий вычислительной сети. Далее представлено описание системы TriadBalance – мультиагентной системы динамической балансировки распределенной имитационной модели.

Описание TriadBalance

Система TriadBalance является управляемой, так как предоставляет возможность настройки правил балансировки. Система является динамической, т.е. балансировка

выполняется прямо во время процесса имитации. Это позволяет своевременно реагировать на всевозможные изменения поведения моделируемой системы и в случае необходимости выполнить перераспределение объектов имитационной модели. Для системы имитации Triad.Net данная особенность обязательна, так как система позволяет изменять имитируемую модель прямо во время процесса имитации, и заранее нельзя точно знать, какие нагрузки выпадут на тот или иной узел или канал сети. Иначе возможно возникновение тупикового состояния системы, например, если канал будет перегружен, то это приведет к перезапуску системы, следовательно, к неполному результату процесса имитации и большим дополнительным расходам.

Система TriadBalance является мультиагентной, т.е. система состоит из множества агентов, выполняющих определенные для них задачи, и вместе реализующие балансировку распределенной имитационной модели. Это позволяет:

- распределить нагрузку системы на узлы сети, так как агенты работают на разных узлах сети;
- ускорить выполнение балансировки, так как агенты, взаимодействующие с определенными частями системы имитации, находятся в тех же узлах сети, что и обрабатываемые ими части;
- унифицировать систему балансировки для различных систем имитации, т.е. при переносе системы балансировки на другую систему имитации (или при изменении данной) появляется возможность переписывать не полностью всю систему, а только некоторых ее агентов (например, непосредственно работающих с системой имитации). Это позволяет распараллелить процесс разработки системы и облегчить процесс ее сопровождения.

Данные свойства проявляют себя особенно сильно при реализации крупных имитационных моделей, с большим числом объектов, и широкой сетью, в которой выполняется имитация.

Описание процесса балансировки

Перед запуском системы проводится предварительная статическая балансировка модели, т.е. определяется начальное расположение объектов модели на вычислительных узлах сети.

Хотя пользователь знает, как будет вести себя система в тот или иной период времени моделирования, он, скорее всего, не сможет однозначно определить постоянное выгодное расположение объектов. Поэтому следующим шагом является настройка системы балансировки. Пользователь, основываясь на знаниях о модели (он знает, как должна работать модель), модифицирует правила балансировки. На основании данных правил агенты будут принимать решение о переносе объектов модели с одного вычислительного узла сети на другой.

После внесения правил балансировки пользователь запускает систему балансировки на каждом узле сети, которая начинает сам процесс балансировки.

После запуска системы TriadBalance агент анализа начинает опрашивать систему сбора статистики для получения данных о состоянии соответствующего вычислительного узла, а также собирает информацию через информационные процедуры системы имитации Triad.Net, для получения данных о состоянии модели на данном вычислительном узле (рис. 1). Получив необходимую информацию, агент анализа начинает исследовать ее на основании правил балансировки введенных пользователем. Исследование правил построено на принципах работы экспертных систем и баз знаний.

В результате проведенного исследования агенты анализа могут получить два возможных итога – перераспределение объектов не требуется или требуется. В первом случае агент анализа снова запрашивает информацию о системе и заново повторяет

анализ системы. Во втором случае агент анализа сообщает агенту распределения о принятом решении, который, в свою очередь, на основании пользовательских правил принимает решение о том, какие именно объекты модели необходимо переносить, а также выбирает целевые вычислительные узлы сети.

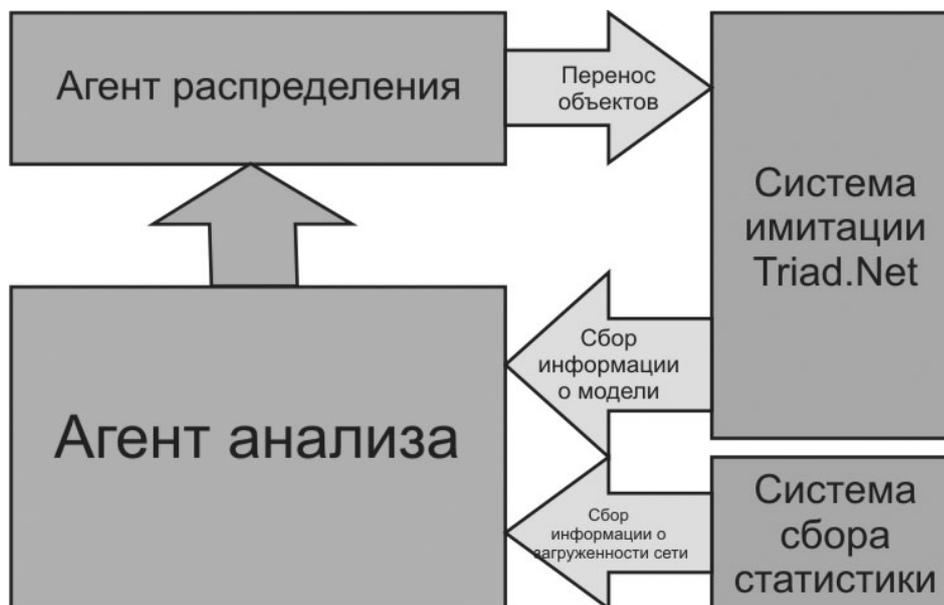


Рис. 1. Структура системы балансировки

Приняв соответствующее решение, агенты распределения, используя разработанные волновые алгоритмы, синхронизируются. После синхронизации они останавливают процесс моделирования, отправив сообщение системе имитации о необходимости остановки процесса. Далее они запрашивают необходимые объекты у системы (в виде потоков сериализованных данных) и передают полученные объекты агентам распределения узлов, на которые необходимо переместить эти объекты. Получив объект, агент распределения отдает его системе имитации для включения его в данный вычислительный узел. После завершения агенты распределения, синхронизируясь, запускают систему. После этого повторяется сбор информации о системе, на основании которой снова проводится анализ для решения задачи о перераспределении объектов системы. Синхронизация и взаимодействие агентов расположенных на разных узлах сети выполняется с использованием волновых алгоритмов.

Реализация агента анализа

Агент анализа собирает информацию о системе в целом (т.е. о состоянии вычислительных узлов и каналов связи между ними, а также о состоянии системы имитации) через определенные интервалы времени. После этого ему необходимо решить, требуется или нет перераспределение объектов системы имитации на вычислительных узлах сети. Принятие данного решения можно разбить на следующие этапы.

А. Этап анализа пользовательских правил. На этом этапе агент анализа считывает предварительно подготовленные пользователем правила их XML-файла. При считывании правил из файла агент выполняет следующий алгоритм.

1. Считывает текущее правило и получает значения его параметров.
2. Используя технологию .Net Reflection, создает объект, соответствующий типу метаправила.
`string xnode_meta = xnode.Attributes["meta"].Value;`

```
Type xnode_type = Type.GetType(xnode_meta);  
MetaRule xnode_metarule=(MetaRule)Activator.CreateInstance(xnode_type);
```

3. Используя технологию .Net Reflection, проверяет соответствие параметров правила параметрам соответствующего типа метаправила.
4. Добавляет полученное правило в коллекцию правил.
5. Переходит к следующему правилу и повторяет алгоритм.

После выполнения описанной процедуры переходим к следующему этапу.

В. Этап построения внутреннего представления о системе в целом. Агент анализа обладает собственным представлением о строении системы имитации и сети вычислительных узлов. Данное представление создается во время процесса анализа. Оно состоит из представления о вычислительных узлах сети и из представления об объектах имитационной модели, а также из множества связей между вычислительными узлами сети, между объектами модели имитации и между объектами и узлами.

При этом агент имеет представление о свойствах каждого из известных ему элементов. Например, агент имеет представление о загруженности вычислительных узлов, о загрузке вносимой каждым объектом модели имитации, о пропускной способности каналов связи между вычислительными узлами сети, о временной последовательности взаимодействий между объектами системы.

Построение внутреннего представления происходит на различных этапах работы агента анализа, в течение всего времени процесса балансировки. Сначала агент строит представление о собственном узле и объектах, на нем располагающихся (рис. 2).

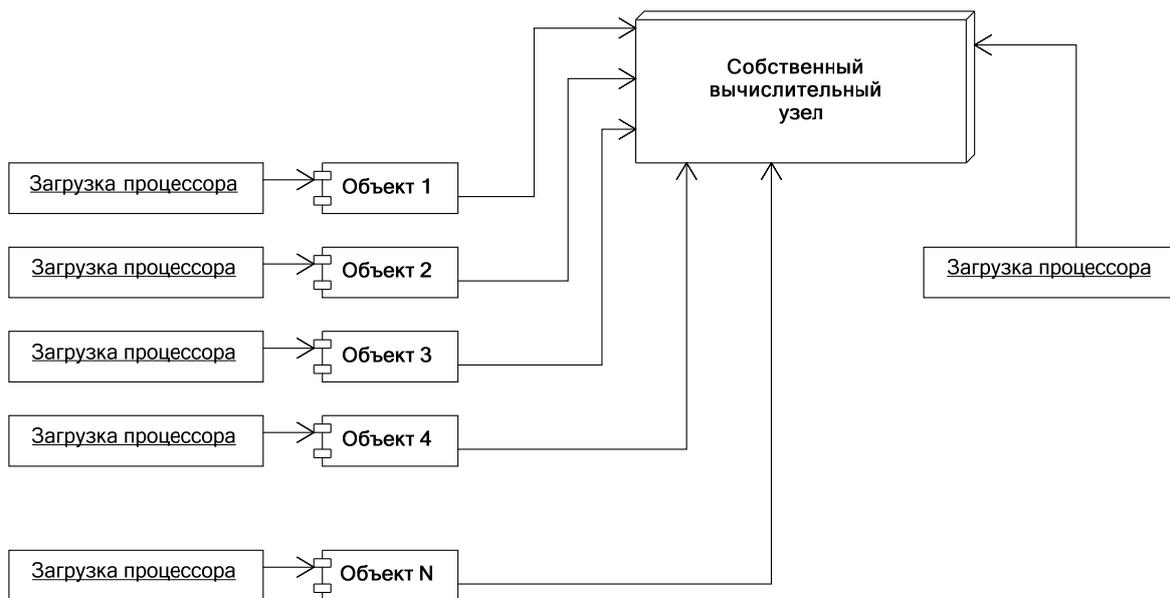


Рис. 2. Начальное представление агента о системе

Далее, во время анализа правил, агент получает знания о существовании других вычислительных узлов и объектах модели имитации (рис. 3).

Новые знания об окружающей его системе он также получает от агента распределения, который сообщает ему об объектах переносимых с других узлов и на другие узлы (рис. 4).

Таким образом, в течение всего процесса балансировки у агента анализа расширяется представление о системе, и, следовательно, он начинает по-другому анализировать складывающиеся ситуации.

После процесса построения (или изменения) внутреннего представления о системе агент переходит к следующему этапу.

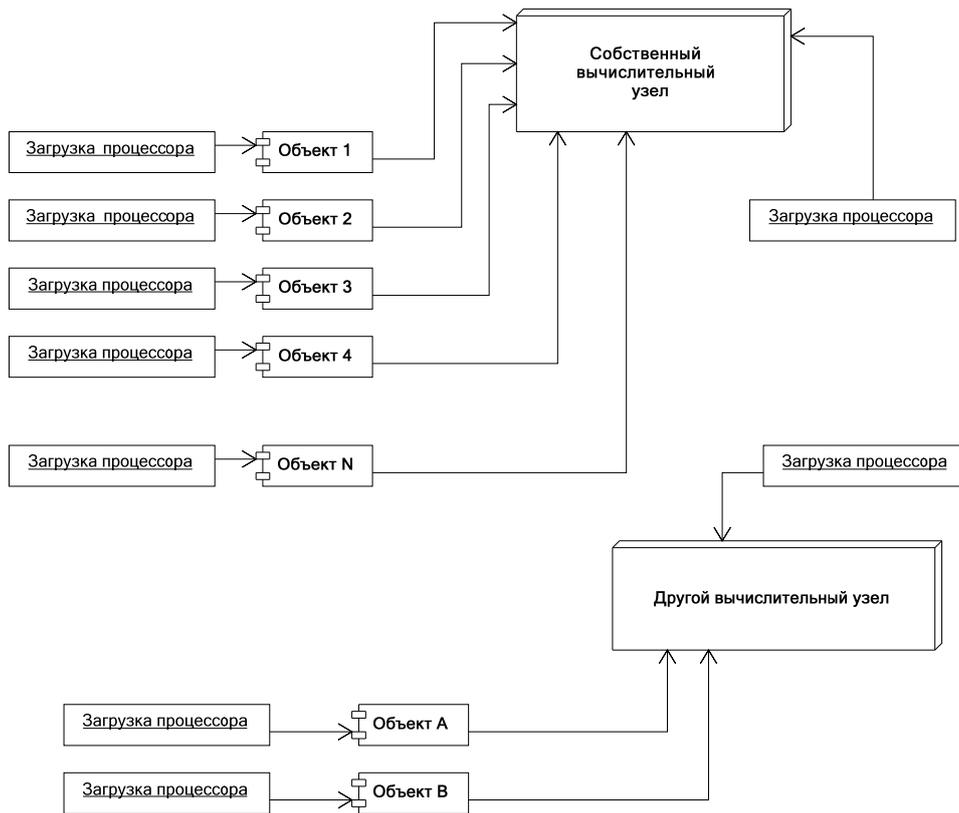


Рис. 3. Новое представление агента о системе

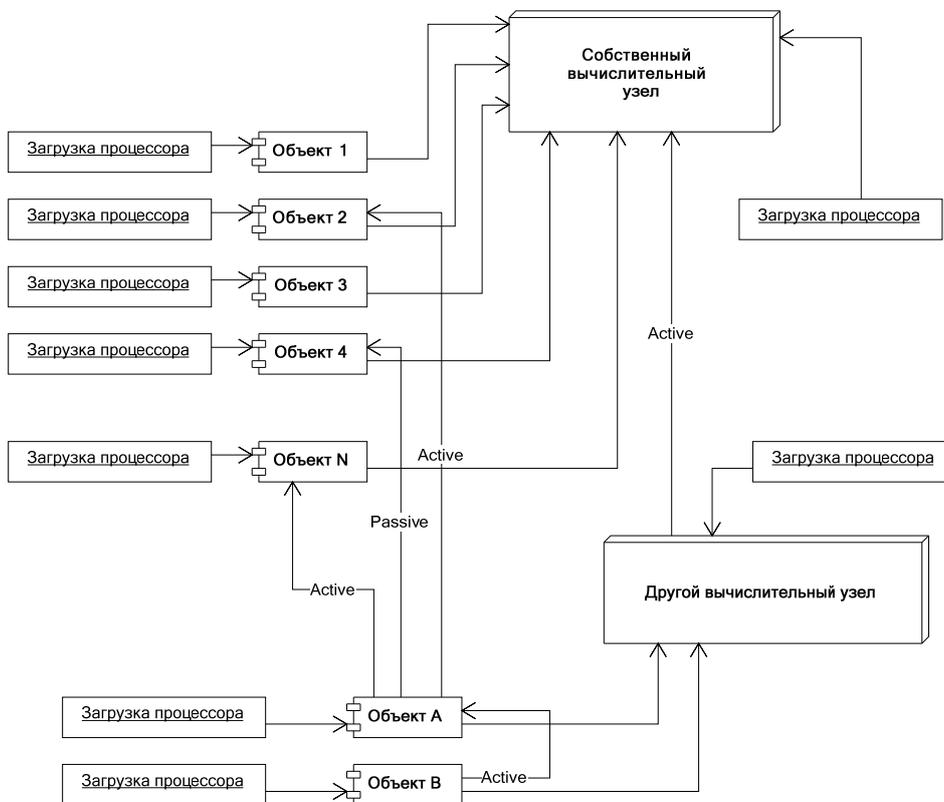


Рис. 4. Окончательное представление агента о системе

С. Этап анализа системы на основе правил балансировки. Агент анализа переходит к этапу анализа системы. Задача этого этапа – на основе правил и внутреннего представления системы определить, необходимо ли выполнять перераспределение объектов модели имитации. Все правила имеют определенную структуру, описанную в метаправилах системы, определяющих также и способ обработки правил.

Приведем пример: в системе описано метаправило с форматом «IF LOADING @MachineName MORE @Range THEN DISTRIBUTE FROM @MachineName», и параметрами @MachineName и @Range, которое означает, что если загрузка процессора на узле @MachineName больше значения @Range, то необходимо перераспределить часть объектов с вычислительного узла @MachineName на другие вычислительные узлы. Пользователь на основе данного метаправила создает правило балансировки «IF LOADING MyMachineName MORE 60 THEN DISTRIBUTE FROM MyMachineName».

Данный набор метаправил – общий и для агента анализа, и для агента распределения, только первый из данных правил извлекает знания о необходимости переноса объектов имитационной модели вообще, а второй – о конкретных переносимых объектах модели и целевых вычислительных узлах. Также агент анализа на основе этих правил извлекает знания для внутреннего представления системы, например, он «узнает» о других вычислительных узлах и о других объектах, которые описаны в правилах.

Волновой алгоритм синхронизации агентов распределения

Перед началом перераспределения агентов в сети необходимо выбрать одного из агентов распределения, который за это будет отвечать. Данный алгоритм относится к классу волновых алгоритмов выбора лидера [4]. Для выбора лидера среди агентов распределения не подходят известные алгоритмы, поэтому нами был разработан алгоритм выбора подходящий для разрабатываемой системы.

Данный алгоритм основан на предположении, что к началу выполнения алгоритма все процессы находятся в одном состоянии, т.е. все агенты распределения готовы выполнить остановку системы.

Каждый агент посылает всем своим соседям сообщение о том, что он хочет быть лидером. Сообщение должно содержать время отправки MessageSignalTime (вплоть до миллисекунд), идентификатор StarterIdent агента распределения (таким образом, каждый агент распределения должен иметь свой собственный уникальный для сети идентификатор DistributionAgentIdentifier), который изначально послал это сообщение, и идентификатор LastIdent агента, который переслал это сообщение. Получая такие сообщения, агент распределения сравнивает время отправки своего сообщения Time со временем MessageSignalTime отправки сообщения, которое пришло на его узел от другого агента распределения. Если значение Time больше значения MessageSignalTime, то агент становится проигравшим. Если значения равны, то сравниваются идентификаторы DistributionAgentIdentifier и StarterIdent, если первое значение больше, то агент становится проигравшим. Если после этих сравнений агент не стал проигравшим, то он увеличивает счетчик Counter на единицу. Если значение Counter стало равно DistributionAgentsCount (общее количество агентов распределения в сети, которое должен знать каждый агент распределения) минус один и при этом агент еще не проигравший, то агент становится выигравшим. После этого агент устанавливает свой идентификатор в поле LastIdent сообщения и пересылает его всем своим соседям (кроме тех от которого сообщение пришло и от которого сообщение отправлено).

```
Const D: integer = количество агентов в сети;  
Var DistributionAgentIdent: string = идентификатор агента;  
Computers – все соседи агента;  
Messages – полученные сообщения;
```

State – текущее состояние;
Time – время отправки сообщения;
Counter – счетчик полученных сообщений.

```
Procedure StartSending() – отправка сообщения
Var Message – сообщение
Begin
  Message.MessageSignalTime = Time.Now; //Текущее время
  Message.StarterIdent = DistributionAgentIdent;
  Message.LastIdent = DistributionAgentIdent;
  Foreach (CurrentComputer Computers)
  Begin
    SendMessage(Message, CurrentComputers); //Отправка сообщения
  End;
End;
Procedure GiveMessage(message) – прием сообщения
Var
Begin
  If (message Messages)
  Begin
    If (message.MessageSignalTime < Time)
    State = States.Lost;
    Else if (message.MessageSignalTime = Time и DistributionAgentIdent > mes-
    sage.StarterIdent)
    State = States.Lost;
    Counter = Counter + 1;
    Messages.Add(message);
    Foreach (CurrentComputer Computers \ {messages.StarterIdent, message.LastIdent})
    Begin
      SendMessage(Message, CurrentComputers); //Отправка сообщения
    End;
  End;
End;
```

Выигравший агент распределения посылает сообщение останова системы, после подтверждения которого он рассылает сообщение всем своим соседям о том, что можно начать перенос объектов, а те, в свою очередь, пересылают его всем своим соседям. Далее выполняется перенос объектов с незагруженного. После завершения переноса каждый из проигравших агентов посылает сообщение лидеру о том, что он завершил перенос. Получив DistributionAgentsCount – 1 таких сообщений, лидер сообщает системе о том, что необходимо запустить ее. На этом заканчивается перенос объектов системы.

Заключение

В работе представлен мультиагентный подход к решению задачи балансировки распределенной имитационной модели. В мультиагентной системе балансировки TriadBalance выделено два типа когнитивных агентов: агент анализа вычислительных узлов сети и модели имитации, агент распределения. Этот подход дает возможность использовать систему балансировки для самых разнообразных имитационных моделей, не привязываясь только к определенному их классу, так как позволяет настраивать

процесс балансировки на конкретную имитационную модель и сеть, в которой происходит процесс моделирования.

На данном этапе была выполнена следующая работа: изучены различные подходы к решению задачи балансировки, разработана структура и архитектура системы балансировки, определены метаправила процесса балансировки, разработано внутреннее представление о системе имитации, подробно разработана архитектура агента анализа, разработан и реализован волновой алгоритм синхронизации агентов распределения, реализована работа агента анализа, с частичным построением внутреннего представления системы.

В дальнейшем предстоит:

- разработать волновой алгоритм для синхронизации агентов анализа;
- реализовать полное построение представления о системе имитации;
- определить способ организации когнитивного агента распределения;
- реализовать агента распределения;
- определить модель взаимодействия агентов с внешними системами: подсистемой сбора статистики и системой имитации Triad.Net.
- реализовать взаимодействие с внешними системами.

Литература

1. Миков А.И., Замятина Е.Б., Фатыхов А.Х. Система оперирования распределенными имитационными моделями сетей телекоммуникаций. Труды Второй Всероссийской научной конференции «Методы и средства обработки информации». М.: Изд-во МГУ, 2003 г.
2. Миков А.И., Замятина Е.Б., Осмехин К.А. Метод динамической балансировки процессов имитационного моделирования. / В кн. «Материалы Всероссийской научно-технической конференции «Методы и средства обработки информации МСО-2005», Москва, 3 октября 2005 г.
3. Cao Jiannong, Sun Yudong, Xianbin Wang and Sajal K. Das «Scalable load balancing on distributed web servers using mobile agents», Department of Computer Science and Engineering, The university of Texas at Arlington, Arlington, TX 76019-0015, USA 14 May, 2003.
4. Tel G. Introduction to Distributed Algorithms. Cambridge University Press, 1994.

ОБРАЗОВАТЕЛЬНАЯ ПРОГРАМА КРАТКОСРОЧНОГО ПОВЫШЕНИЯ КВАЛИФИКАЦИИ В ОБЛАСТИ СУПЕРКОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ

В.О. Лукашук, С.Ю. Лукашук

(Уфимский государственный авиационный технический университет)

Научный руководитель – д.ф.-м.н., профессор Р.К. Газизов

(Уфимский государственный авиационный технический университет)

В работе приводится структура образовательной программы краткосрочного повышения квалификации «Основы суперкомпьютерных технологий», разработанной в Уфимском государственном авиационном техническом университете в рамках национального проекта «Образование», методика ее преподавания и результаты апробации.

Введение

Многопроцессорные вычислительные системы (МВС) получают все более широкое распространение в самых различных областях человеческой деятельности. Если пятнадцать-двадцать лет назад такие системы использовались преимущественно в области компьютерного моделирования для изучения сложных физических процессов и для создания высокотехнологичной техники, то в последние годы наблюдается активное внедрение многопроцессорных систем в производственную сферу, медицину, образование и даже средний и малый бизнес. Причины этого достаточно очевидны: прогресс в области электронной техники и информационных технологий привел к существенному снижению стоимости таких систем.

Совсем недавно произошел еще один качественный скачок – появились многоядерные процессоры общего назначения. В результате обычный персональный компьютер превратился в параллельную вычислительную систему. Конечно, количество ядер пока не велико (не превышает четырех), однако прогресс не стоит на месте. Важным является другое – эффективное использование многоядерных систем невозможно без применения методов и инструментов параллельного программирования, которые первоначально разрабатывались для суперкомпьютерных систем. Мир персональных компьютеров постепенно становится параллельным. В связи с этим возникает необходимость подготовки квалифицированных специалистов в области использования многопроцессорных вычислительных систем.

В 2007 г. в рамках Приоритетного национального проекта «Образование» Уфимским государственным авиационным техническим университетом (УГАТУ) был приобретен новый суперкомпьютер, построенный на базе фирменных решений IBM BladeServer. Одновременно с этим остро встал вопрос о переподготовке преподавательского состава университета в области высокопроизводительных вычислений. В качестве первого этапа решения этой задачи в рамках инновационной образовательной программы подготовки кадров в области информационных технологий проектирования, производства и эксплуатации сложных технических объектов преподавателями кафедры высокопроизводительных вычислительных технологий и систем (ВВТиС) была разработана образовательная программа краткосрочного повышения квалификации «Основы суперкомпьютерных технологий». По этой программе в течение двух месяцев была проведена переподготовка более пятидесяти преподавателей и сотрудников университета.

В данной работе кратко описываются структура программы и методика ее преподавания, а также результаты ее апробации.

Цель и задачи образовательной программы

В основу образовательной программы «Основы суперкомпьютерных технологий» был положен более чем семилетний опыт сотрудников кафедры ВВТиС по преподаванию дисциплин, связанных с параллельными вычислениями, для студентов дневной формы обучения [1–3]. На протяжении этого времени подготовка осуществлялась в двух основных направлениях:

- (1) в области организации распределенных вычислительных сред и уделенного доступа к ним;
- (2) в области постановки и решения естественнонаучных задач с использованием МВС.

В рамках первого направления проводится обучение студентов старших курсов специальности 230101 – Вычислительные машины, комплексы, системы и сети по дисциплине «Параллельные алгоритмы и системы». Целью подготовки является обучение высококвалифицированных системных и проблемных программистов для технического и программного обеспечения МВС, подготовки кадров по созданию телекоммуникационных сред высокопроизводительных технологий и подготовка персонала по обслуживанию локальных и глобальных вычислительных сетей.

Второе направление предполагает подготовку специалистов в области математического моделирования с использованием МВС. В УГАТУ подготовка в данной области проводится для студентов специальности 230401 – Прикладная математика и 010501 – Прикладная математика и информатика, поскольку соответствующие государственные стандарты наиболее полно отвечают требованиям, предъявляемым к уровню подготовки специалистов в названной области. Программа подготовки включает изучение студентами таких дисциплин, как «Многопроцессорные системы и параллельное программирование», «Средства параллельного программирования», «Системное программное обеспечение многопроцессорных вычислительных систем» и ряд других.

В связи с приобретением УГАТУ нового суперкомпьютера появилась необходимость в освоении преподавателями и сотрудниками университета основ суперкомпьютерных вычислений для использования новых вычислительных мощностей в научных исследованиях и учебном процессе. Новый суперкомпьютер (рис. 1) с предварительными характеристиками входит в 30-ю редакцию списка TOP500 самых высокопроизводительных компьютеров мира [4].

Параметры системы:

- 266 двухпроцессорных узлов, 532 четырехядерных процессора Intel Xeon 5300 2.33 ГГц;
- полная оперативная память 2.15 ТВ;
- дисковая память 26.7 ТВ;
- ленточная библиотека 8.8 ТВ;
- вычислительная коммуникационная среда – Infiniband;
- операционная система: RedHat Enterprise Linux 4.4;
- потребляемая мощность 95 КВт.

Суперкомпьютер построен на базе фирменных решений IBM BladeServer. Пиковая производительность – 19.8 TFlops, производительность по тесту LINPACK – 15.8 TFlops.

Очевидно, что для эффективного использования такой техники необходимы практические навыки использования МВС. На решение этой задачи и ориентирована, в первую очередь, разработанная образовательная программа.

Принципиальной отличительной особенностью разработанной образовательной программы «Основы суперкомпьютерных технологий» является то, что она ориентирована на слушателей, не имеющих специальной подготовки в области численных методов, методов анализа алгоритмов, системного программирования. В ней сделана попытка комплексно и с единых позиций осветить последние достижения и проблемы в

области современных высокопроизводительных вычислительных технологий. Программа охватывает основные направления развития современной высокопроизводительной вычислительной техники – от суперкомпьютеров и кластерных систем до персональных многоядерных компьютеров. Последним уделено особое внимание, поскольку такие системы получили распространение лишь в течение последних двух лет, и учебные издания по данному направлению практически отсутствуют. Разработанный образовательный продукт частично решил эту проблему.



Рис. 1. Суперкомпьютер УГАТУ (Уфа, 2007 г.)

Целью образовательной программы является приобретение слушателями базовых навыков работы с МВС и создания параллельных вычислительных программ для многопроцессорных и многоядерных систем.

Задачи образовательной программы сводятся к рассмотрению основных архитектурных принципов реализации параллельной обработки данных и методов конструирования параллельных алгоритмов, освоению слушателями технологий разработки параллельных приложений для многопроцессорных систем с общей и распределенной памятью, а также использованию ими современных библиотек параллельных подпрограмм.

Структура образовательной программы и методика преподавания

Для учебно-методического обеспечения образовательной программы был создан учебно-методический комплекс, включающий

- (1) учебную программу дисциплины;
- (2) набор презентаций лекционного курса;
- (3) набор презентаций лабораторного практикума;
- (4) учебно-методическое пособие по методике изучения материалов курса;
- (5) учебно-методическое пособие по методике проведения лекционных занятий и лабораторного практикума;
- (6) контрольные задания в тестовой форме.

Структура образовательной программы приведена в таблице.

Лекционный курс подкреплён набором презентаций (более 500 слайдов), которые наглядно иллюстрируют излагаемый материал, позволяют экономить время лектора на его изложение и способствуют формированию вопросов у аудитории в процессе лекции.

№ раздела	Название раздела	Лекции, час.	Лабораторные работы, час	Всего, час
1	Введение в параллельную обработку данных	3	–	3
2	Виды параллельных вычислительных систем	3	–	3
3	Архитектурные принципы параллелизма	4	–	4
4	Системное программное обеспечение многопроцессорных вычислительных систем	4	–	4
5	Парадигмы параллельного программирования	2	–	2
6	Средства параллельного программирования	2	–	2
7	Средства разработки параллельного программного обеспечения для многоядерных систем	8	16	24
8	Средства разработки параллельного программного обеспечения для кластерных систем	8	8	16
9	Параллельные алгоритмы	4		4
10	Библиотеки параллельных подпрограмм	2	8	10
11	Основы программирования на языке C/C++ (дополнительный раздел)	(4)	(4)	(8)
	Итого	40 (44)	32 (36)	72 (80)

Таблица. Содержание дисциплины

Лабораторный практикум охватывает вопросы создания параллельных программ как для многоядерных и многопроцессорных систем с общей памятью, так и для кластерных систем с использованием современных средств параллельного программирования и библиотек параллельных программ. В рамках курса были разработаны и проведены следующие лабораторные работы:

- (1) написание простейшей параллельной программы на примере вычисления определенного интеграла для многоядерной системы средствами OpenMP (4 часа);
- (2) написание простейшей параллельной программы на примере вычисления суммы числового ряда для кластерной системы средствами MPI (4 часа);
- (3) написание программы численного решения системы линейных уравнений на многоядерной вычислительной системе с использованием библиотеки Intel MathKernel Library (8 часов);
- (4) написание параллельной программы умножения матриц для многоядерной системы средствами WinAPI и OpenMP (8 часов);
- (5) написание параллельной программы умножения матриц для кластерной системы с использованием библиотеки ScaLAPACK (8 часов);
- (6) (дополнительная) создание простейших программ на C/C++ с использованием управляющих структур, массивов и указателей (4 часа).

Для проведения лабораторных работ использовались дисплейный класс кафедры ВВТиС и дисплейный класс кафедры математики. В качестве системы с распределенной памятью использовался учебный Alpha-кластер кафедры ВВТиС (12×Alpha 21164

EV5 / 256 MB / Fast Ethernet 100 Mbit/s). Кластер работает под управлением операционной системы Debian Linux. Несмотря на то, что его пиковая производительность (порядка 12 GFlops) меньше производительности современного двухядерного процессора, он позволяет демонстрировать все особенности создания параллельных приложений для систем с распределенной памятью.

Создание учебных параллельных программ в рамках лабораторного практикума организовано в несколько этапов, в зависимости от изучаемых средств. При выполнении лабораторных работ (1), (3) и (4) в силу хорошей оснащенности дисплейных классов (компьютеры с двухядерными процессорами семейства AMD Athlon 64 X2) написание и отладка программ производится на персональном компьютере. Основным результатом выполнения этих лабораторных работ является грамотное использование механизмов синхронизации потоков и оптимизация работы программы с кэш-памятью.

Для лабораторных работ (2) и (5) разработка программ выполняется в два этапа. На первом этапе производится написание и первоначальная отладка программного кода. Этот этап выполняется на двухядерных однопроцессорных персональных компьютерах, работающих под управлением ОС семейства Linux. Для написания программы используются компиляторы Fortran и C, а распараллеливание осуществляется средствами MPI. Для отладки механизмов обмена данными между отдельными процессами программы на данном этапе используется режим эмуляции, поддерживаемый MPI. Кроме того, наличие многоядерности позволяет проверить ускорение в работе программы при использовании двух ядер по сравнению с одноядерным режимом. Если ускорение отсутствует или оказывается слишком малым, производится переработка программного кода или модификация параллельного алгоритма. На втором этапе параллельная программа запускается на кластере в режиме удаленного доступа. На этом этапе производится окончательная отладка программы и детальное исследование ее эффективности в зависимости от задействованных узлов системы.

Описанная двухступенчатая технология создания параллельных программ хорошо зарекомендовала себя на практике. Она позволила практически полностью исключить кластерные системы из процесса отладки программ, тем самым освободив их для выполнения более важных вычислительных задач.

В конце обучения предусмотрен зачет в тестовой форме. Разработан комплект контрольных заданий в тестовой форме из 25 заданий, каждое задание состоит из 20 тестовых вопросов (по 2 вопроса на каждую тему, изучаемую в дисциплине). Все вопросы считаются равнозначными и не имеют градации по степени сложности. Каждый тестовый вопрос содержит четыре варианта ответов с одним или двумя (указывается в вопросе) правильными. Для успешного прохождения теста необходимо дать правильные ответы не менее чем на 15 вопросов из 20. При этом также выставляются оценки по четырехбалльной шкале по следующим критериям: оценка «отлично» ставится за 19–20 правильных ответов, оценка «хорошо» – за 17–18, оценка «удовлетворительно» – за 15–16 и «неудовлетворительно» – в остальных случаях.

Статистические результаты апробации образовательной программы

В ноябре–декабре 2007 г. курс повышения квалификации по образовательной программе «Основы суперкомпьютерных технологий» прошли 50 преподавателей и сотрудников университета, в том числе два профессора, 12 доцентов, восемь старших преподавателей, 18 ассистентов, 10 научных сотрудников и программистов университета. По итогам прохождения курсов повышения квалификации был проведен зачет в тестовой форме, результаты которого представлены на рис. 2. Видно, что 90% обучаемых успешно освоили материал, причем более 70% выполнили задание теста на «хорошо» и «отлично».

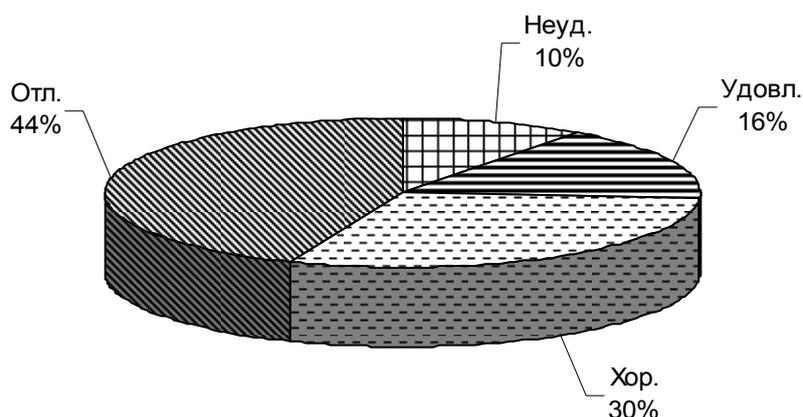


Рис. 2. Результаты зачета по дисциплине

Большую часть обучаемых составляли аспиранты и молодые ученые (74%); распределение оценок в этой группе иллюстрирует рис. 3. Видно, что процент успешно освоивших материал в этой группе оказывается несколько выше аналогичного показателя по всему контингенту обучаемых (рис. 2), однако процент освоивших материал на «хорошо» и «отлично» остался на прежнем уровне.

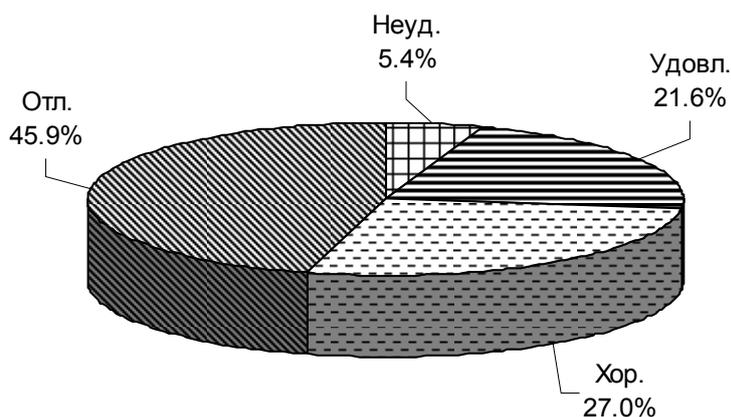


Рис. 3. Результаты зачета среди аспирантов и молодых ученых

Приведенные результаты апробации образовательной программы подтверждают эффективность выбранной методики ее преподавания, позволяющей успешно освоить основы суперкомпьютерных технологий людям, прежде не знакомым с данной предметной областью.

Заключение

В работе приведено краткое описание структуры образовательной программы краткосрочного повышения квалификации «Основы суперкомпьютерных технологий» и методики ее преподавания. Приведенные статистические результаты ее апробации на преподавателях и научных сотрудниках университета свидетельствуют о том, что за короткое время основами суперкомпьютерных технологий может успешно овладеть человек, не имеющий до этого никаких представлений об этой области. По материалам программы разработаны и готовятся к изданию учебное пособие с грифом УМО, лабораторный практикум, а также ряд учебно-методических пособий. Отметим, что приведенная программа является лишь вводной строкой в мир параллельных вычислений и будет иметь свое продолжение в ряде специальных образовательных программ.

Литература

1. Кусимов С.Т., Газизов Р.К., Ковалева (Лукашук) В.О., Лукашук С.Ю. и др. Научно-методическое и научное обеспечение подготовки кадров в Башкирском региональном центре высокопроизводительных вычислений. / Сводный отчет по проекту ГР 01200311026 ФЦП “Интеграция” № 02200305050. – Уфа, 2002. – 191 с.
2. Гузаиров М.Б., Газизов Р.К., Лукашук С.Ю., Хисамутдинов Р.А. Высокопроизводительные вычисления в УГАТУ: от подготовки кадров к научным исследованиям // Материалы научно-технического совещания «Современные наукоемкие технологии в промышленности России: высокопроизводительные вычисления и CALS-технологии», Уфа, 25–28 октября 2004 г. – Уфа, 2005. – С. 93–98.
3. Газизов Р.К., Лукашук С.Ю., Тулебаев С.Д. Основы параллельного программирования с использованием MPI: Уч. пос. с грифом УМО. – Уфа: Изд-во УГАТУ, 2004. – 90 с.
4. TOP 500. Supercomputer sites. – Режим доступа: <http://www.top500.org>

РЕАЛИЗАЦИЯ КВАНТОВОГО ЛОГИЧЕСКОГО ВЕНТИЛЯ CCNOT НА ОСНОВЕ ЗЕЕМАНОВСКОЙ СТРУКТУРЫ УРОВНЕЙ В АТОМАХ РУБИДИЯ

А.И. Трифанов

Научные руководители – д.ф.-м.н., профессор И.Ю. Попов,
д.ф.-м.н., профессор Г.П. Мирошниченко

В работе строится математическая модель универсального квантового логического вентиля CCNOT на элементной базе квантового оптического компьютера. Описывается схема построения устройства, использующая эффекты квантовой оптики. Вентиль реализуется на основе ячейки с парами рубидия при низкой температуре, помещенной в сильное магнитное поле. Роль кубитов играет поляризация пучков света, падающих на ячейку (в другом варианте – однофотонное состояние света).

1. Введение

В настоящее время особый интерес представляют работы по созданию квантового компьютера. Квантовые эффекты перепутывания и когерентности делают эту область привлекательной при построении телекоммуникационных систем и сетей, а также при решении задач, которые в классическом компьютере требуют экспоненциального времени расчета, за полиномиальное время [1, 2]. Работы в данной области также востребованы в новой области, называемой квантовой криптографией, которая позволяет создать сверхустойчивые протоколы передачи секретных сообщений.

Квантовой единицей информации служит квантовый бит (кубит), который, в отличие от классического бита, может находиться в суперпозиции состояний нуля и единицы. Благодаря этому достигается параллелизм, которого невозможно достичь в классических компьютерах.

Главная проблема при разработке квантового компьютера – это вопрос создания элементной базы. Логический элемент, осуществляющий преобразование кубита или системы кубитов, называется квантовым вентилем. В настоящее время предложено множество вариантов реализации таких устройств. Однако математическое моделирование вентиля, преобразующего систему кубитов, – очень сложная задача, тем сложнее, чем больше система. Также одной из проблем в этой области является выбор интерпретации кубита, т.е. физической величины, используемой для кодирования информации. Наиболее известные – это спин электрона, ядерный спин, поляризация фотона (однофотонное состояние) [3].

В настоящей работе предложена реализация квантового логического устройства CCNOT (дважды контролируемое НЕ), которое осуществляет преобразование трехкубитового состояния в соответствии с таблицей (см. табл. 1). Данное логическое устройство является универсальным в том смысле, что одних таких устройств достаточно для выполнения любой квантовой логической операции [2, 4].

Описанное преобразование соответствует следующей матрице:

$$Q_{8 \times 8} = \begin{cases} \text{diag} \{ \delta_{i,i} \}_{i=1}^7 \\ Q_{78} = Q_{87} = 1, Q_{8,8} = 0 \end{cases} \quad (1.1)$$

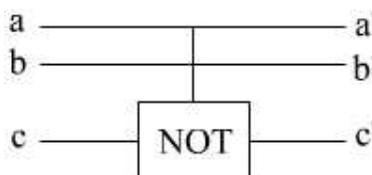


Рис. 1. Логический вентиль CCNOT

abc	a'b'c'
000	000
001	001
010	010
011	011
100	100
101	101
110	111
111	110

Таблица 1. Преобразование CCNOT

Информация кодируется в поляризации фотона. Данный выбор обоснован тем, что фотон сам по себе является идеальным носителем информации, так как не подвержен эффектам декогерентизации, разрушающим квантовые вычисления. Однако имеется и свой недостаток – фотоны слишком слабо взаимодействуют друг с другом, чтобы построить перепутанное состояние нескольких кубитов. Преодолеть данное затруднение можно, создав специальные условия в среде, называемые вынужденной электромагнитной прозрачностью.

В качестве среды распространения, следуя [5, 6], выбирается ячейка с атомами ^{87}Rb , помещенная в сильное магнитное поле. В веществе выбирается система зеемановских уровней, обладающая следующим свойством: переходы между уровнями соответствуют строго определенному виду круговой поляризации фотона. В зависимости от того, сколько пучков в среде распространяется, набег фазы поляризации контролируемого пучка под действием гигантского эффекта Керра меняется. Построенный таким образом контролируемый фазовый сдвиг является основой для реализации вентиля (достаточно лишь на результирующее состояние подействовать однокубитовым преобразованием Адамара, что является несложной задачей).

В работе [5] подобная схема исследовалась для реализации квантового двухкубитового вентиля CNOT (контролируемое НЕ). Там же было использовано пять атомных уровней (так называемая М-схема) с двумя полями (двухкубитовое состояние). В данной статье изучается более общая и сложная динамика системы вещество–свет, включающая три поля. Доказывается возможность построения логического вентиля CCNOT на базе предложенной физической системы.

2. Описание работы М-схемы

В данной работе логический вентиль реализован на зеемановских уровнях в атомах рубидия, обладающих тем свойством, что переходы зависят от вида поляризации падающего излучения. На рис. 2 изображена схема уровней, получившая название М-схемы.

Через ячейку проходят три хорошо стабилизированных по частоте слабых пучка света с частотами ω_p , ω_c и ω_t . Пучки с частотами ω_p и ω_c являются контролирующими, а ω_t – контролируемым. Для этих пучков создаются условия взаимодействия посредством сильного эффекта Керра. В зависимости от поляризаций входного излучения схема будет вести себя по-разному.

В начальном состоянии заселен только основной, первый уровень. Переход из основного состояния $|1\rangle$ в состояние $|2\rangle$ осуществляет электромагнитная волна с частотой ω_p и круговой поляризацией по часовой стрелке (σ_p^+), переход из состояния $|2\rangle$ в

$|3\rangle$ – с частотой волны ω_c и поляризацией против часовой стрелки (σ_c^-), а из состояния $|3\rangle$ в состояние $|4\rangle$ переход соответствует частоте ω_t и поляризации против часовой стрелки (σ_t^-). Сильное поле с частотой ω_s между $|4\rangle$ и $|5\rangle$ уровнями введено для того, чтобы согласовывать прохождение этих трех пучков. Символами $\Omega_{ij} = \frac{1}{2} \bar{F}_l \bar{d}_{ij}$, ($i, j = 1..5, i \neq j, l = \{p, c, t, s\}$) обозначены частоты Раби соответствующих переходов, где d_{ij} – дипольный момент этого перехода, F_l – поле, действующее на переходе ij в соответствии с рис. 2.

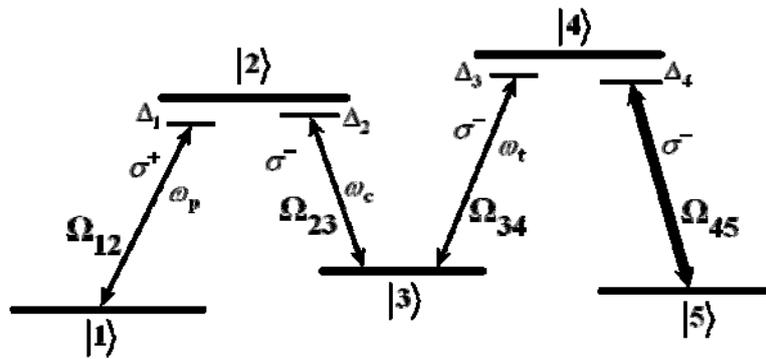


Рис. 2. Конфигурация уровней в М-схеме

Рассмотрим следующую ситуацию: пусть ω_p имеет поляризацию σ_p^- , ω_c – поляризацию σ_c^- и ω_t – поляризацию σ_t^- . При такой комбинации поляризаций нелинейный набег фазы наблюдаться не будет, так как световой пучок ω_p не является резонансным для перехода $|1\rangle \rightarrow |2\rangle$. Световая волна ω_p пройдет через вещество, практически не взаимодействуя с ним (т.е. не будет возбуждать электроны до уровня $|2\rangle$), а значит, взаимодействие и остальных пучков со средой будет пренебрежимо мало. Суммарный нелинейный набег фазы в этом случае будет также пренебрежимо мал.

Рассмотрим другой случай, когда каждый из трех входных пучков находится в резонансе со своим переходом. Все три поля взаимодействуют с веществом и друг с другом. При этом их групповая скорость мала, а время взаимодействия велико. Наблюдается сильный нелинейный фазовый сдвиг поляризации.

кубиты	1 → 2	2 → 3	3 → 4	Суммарный набег фазы $\Delta\varphi$
000	-	-	-	0
001	-	-	-	0
010	-	-	-	0
011	-	-	-	0
100	+	-	-	0
101	+	-	-	0
110	+	+	-	0
111	+	+	+	π

Таблица 2. Наличие переходов между уровнями и суммарный нелинейный набег фазы поляризации в зависимости от поляризации входных пучков

В табл. 2 приведены суммарные набег фаз для возможных распределений поляризации в пучках. В левом столбце единица соответствует круговой поляризации, вызывающей переход между соответствующими уровнями, нуль – перехода нет. Левая

позиция в тройке чисел соответствует пучку с частотой ω_p , средняя – ω_t , правая – ω_c . Во втором, третьем и четвертом столбцах плюс соответствует наличию резонансного перехода между уровнями, минус – его отсутствию.

Матрица, соответствующая описанному преобразованию, имеет вид [7]

$$M = \text{diag} \left\{ 1 - 2\delta_{i,8} \right\}_{i=1}^8 \quad (2.1)$$

Для получения матрицы преобразования CCNOT достаточно совершить преобразование

$$Q = (I \otimes I \otimes H) M (I \otimes I \otimes H), \quad (2.2)$$

где $H = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ – матрица преобразования Адамара, а I – единичная матрица.

Таким образом, можно сказать, что возможна реализация квантового логического устройства CCNOT на базе М-схемы, если будет доказано наличие существенного по сравнению с другими случаями нелинейного набега фазы поляризации в М-схеме. Следующая часть статьи посвящена этому вопросу.

3. Нахождение суммарных нелинейных набегов фаз

Для нахождения суммарного нелинейного набега фазы поляризации $\Delta\varphi = \Delta\varphi_p + \Delta\varphi_c + \Delta\varphi_t$ необходимо найти восприимчивость на каждом переходе атомной схемы [6,7]:

$$\chi_{ij} = \frac{\langle \Psi | i \rangle \langle j | \Psi \rangle}{F_{ij}} d_{ij} \quad (i, j = 1 \dots 5, i \neq j). \quad (3.1)$$

После этого нахождение набегов фаз трудности представлять не будет:

$$\begin{aligned} \Delta\varphi_p &= k_0^p L \sqrt{1 + 4\pi N \chi_{12}}, \\ \Delta\varphi_c &= k_0^c L \sqrt{1 + 4\pi N \chi_{23}}, \\ \Delta\varphi_t &= k_0^t L \sqrt{1 + 4\pi N \chi_{34}}, \end{aligned} \quad (3.2)$$

где k_0^i , $i = \{p, c, t\}$ – волновой вектор соответствующего пучка в вакууме, N – число атомов среды в единице объема, L – длина ячейки, Ψ – волновая функция М-схемы, для нахождения которой будем решать уравнение Шредингера:

$$i \frac{\partial \Psi}{\partial t} = H \Psi. \quad (3.3)$$

Оператор Гамильтона запишем в представлении взаимодействия:

$$H = H_0 + V,$$

где $H_0 = \text{diag} \{ E_i \}_{i=1}^5$ – оператор Гамильтона свободной системы с собственными числами E_i . Оператор взаимодействия V имеет вид:

$$V = \begin{pmatrix} 0 & d_{12}(t) & 0 & 0 & 0 \\ d_{12}^*(t) & 0 & d_{23}(t) & 0 & 0 \\ 0 & d_{23}^*(t) & 0 & d_{34}(t) & 0 \\ 0 & 0 & d_{34}^*(t) & 0 & d_{45}(t) \\ 0 & 0 & 0 & d_{45}^*(t) & 0 \end{pmatrix}. \quad (3.4)$$

Элементы матрицы V , представляют собой коэффициенты взаимодействия [17]:

$$\begin{aligned}
 d_{12}(t) &= \Omega_{12} e^{i\omega_p \hbar t}, & d_{12}^*(t) &= \Omega_{12}^* e^{-i\omega_p \hbar t} \\
 d_{23}(t) &= \Omega_{23} e^{-i\omega_c \hbar t}, & d_{23}^*(t) &= \Omega_{23}^* e^{i\omega_c \hbar t} \\
 d_{34}(t) &= \Omega_{34} e^{i\omega_t \hbar t}, & d_{34}^*(t) &= \Omega_{34}^* e^{-i\omega_t \hbar t} \\
 d_{45}(t) &= \Omega_{45} e^{-i\omega_s \hbar t}, & d_{45}^*(t) &= \Omega_{45}^* e^{i\omega_s \hbar t}
 \end{aligned} \tag{3.5}$$

Далее представляется удобным перейти к резонансному приближению. При этом мы будем работать не с самими уровнями атомной системы (собственными числами оператора H_0), а с отстройками частот переходов между энергетическими уровнями атома и частотами световых пучков, которые связывают данные уровни. Это приводит к тому, что коэффициенты связи перестают зависеть от времени, что существенно упрощает анализ и решение задачи. Мы будем считать, что отстройки частот световых полей от частот переходов между атомарными уровнями много меньше самих переходов между уровнями. Это позволяет нам правомерно строить решение нашей задачи в резонансном приближении.

Теперь, для удобства и корректности дальнейшего изложения, мы определимся с обозначениями частот, а именно: символами ω_p , ω_c и ω_t будем обозначать частоты электромагнитных полей, падающих на ячейку, тогда как символами ω_{ij} , $i, j = 1..5$, $i \neq j$ будут обозначены частоты переходов между энергетическими уровнями в атомах рубидия.

Волновую функцию будем искать в следующем виде:

$$\Psi(t) = \exp(-i\bar{H}_0 t) \tilde{\Psi}(t), \tag{3.6}$$

где \bar{H}_0 – оператор Гамильтона системы, свободной от внешних возмущений, в резонансном приближении:

$$\bar{H}_0 = \text{diag} \{ \bar{E}_{i,i} \}_{i=1}^5. \tag{3.7}$$

Матрица $e^{-i\bar{H}_0 t}$ имеет вид

$$e^{-i\bar{H}_0 t} = \text{diag} \left\{ \exp(-i\bar{E}_j t) \right\}_{j=1}^8$$

и обладает следующими свойствами:

$$e^{-i\bar{H}_0 t} H_0 e^{i\bar{H}_0 t} = H_0, [\bar{H}_0 H_0] = 0.$$

Далее, чтобы написать уравнение Шредингера в резонансном приближении, необходимо найти элементы матрицы \bar{H}_0 . Запишем условия выбранного приближения:

$$\begin{cases}
 \exp[i\omega_p \hbar t] \exp[-i(\bar{E}_2 - \bar{E}_1)t] = 1 \\
 \exp[-i\omega_c \hbar t] \exp[-i(\bar{E}_2 - \bar{E}_3)t] = 1 \\
 \exp[i\omega_t \hbar t] \exp[-i(\bar{E}_4 - \bar{E}_3)t] = 1 \\
 \exp[-i\omega_s \hbar t] \exp[-i(\bar{E}_4 - \bar{E}_5)t] = 1
 \end{cases} \tag{3.8}$$

Решая данную систему относительно \bar{E}_i ($i = 1..5$) с предположением $\bar{E}_1 = E_1$, получаем:

$$\begin{cases} \bar{E}_1 = E_1 \\ \bar{E}_2 = E_1 + \omega_p \hbar \\ \bar{E}_3 = E_1 + \omega_p \hbar - \omega_c \hbar \\ \bar{E}_4 = E_1 + \omega_p \hbar - \omega_c \hbar + \omega_t \hbar \\ \bar{E}_5 = E_1 + \omega_p \hbar - \omega_c \hbar + \omega_t \hbar - \omega_s \hbar \end{cases} \quad (3.9)$$

Преобразуем далее уравнение Шредингера и получим его вид в резонансном приближении. Для этого подставим волновую функцию (3.6) в исходное уравнение (3.3):

$$i\dot{\tilde{\Psi}}(t) = (H_0 - \bar{H}_0)\tilde{\Psi}(t) + \exp(-i\bar{H}_0 t)V \exp(i\bar{H}_0 t)\tilde{\Psi}(t). \quad (3.10)$$

Второе слагаемое в (3.10) представляет собой оператор возмущения в резонансном представлении. Учитывая выражения для коэффициентов $d_{ij}(t)$ и $d_{ij}^*(t)$ (3.5), получим:

$$e^{-i\bar{H}_0 t} V e^{i\bar{H}_0 t} = \begin{pmatrix} 0 & \Omega_{12} & 0 & 0 & 0 \\ \Omega_{12}^* & 0 & \Omega_{23} & 0 & 0 \\ 0 & \Omega_{23}^* & 0 & \Omega_{34} & 0 \\ 0 & 0 & \Omega_{34}^* & 0 & \Omega_{45} \\ 0 & 0 & 0 & \Omega_{45}^* & 0 \end{pmatrix}. \quad (3.11)$$

Оператор в первом слагаемом в (3.10) – разность двух операторов,

$$H_0 - \bar{H}_0 = \text{diag} \{ \Delta_i \}_{i=1}^5,$$

и для удобства введены обозначения $\Delta_i = E_i - \bar{E}_i$, $i = 1..5$.

Функцию $\tilde{\Psi}(t)$ разложим по собственным векторам атомной системы:

$$\tilde{\Psi}(t) = \sum_{j=1}^5 c_j |j\rangle. \quad (3.12)$$

Теперь можно записать получившееся уравнение Шредингера:

$$i \begin{pmatrix} \dot{c}_1(t) \\ \dot{c}_2(t) \\ \dot{c}_3(t) \\ \dot{c}_4(t) \\ \dot{c}_5(t) \end{pmatrix} = \begin{pmatrix} \Delta_1 & \Omega_{12} & 0 & 0 & 0 \\ \Omega_{12}^* & \Delta_2 & \Omega_{23} & 0 & 0 \\ 0 & \Omega_{23}^* & \Delta_3 & \Omega_{34} & 0 \\ 0 & 0 & \Omega_{34}^* & \Delta_4 & \Omega_{45} \\ 0 & 0 & 0 & \Omega_{45}^* & \Delta_5 \end{pmatrix} \begin{pmatrix} c_1(t) \\ c_2(t) \\ c_3(t) \\ c_4(t) \\ c_5(t) \end{pmatrix}, \quad (3.13a)$$

или в векторной форме

$$i\dot{\vec{C}} = H_V \vec{C}. \quad (3.13b)$$

Из условия начальной заселенности первого уровня получаем начальные условия:

$$\vec{C}(0) = (1 \ 0 \ 0 \ 0 \ 0)^T.$$

4. Решение системы уравнений методом стационарной теории возмущений

Будем решать получившуюся систему дифференциальных уравнений методами стационарной теории возмущений. Для этого представим матрицу H_V в виде суммы

$$H_V = H_V^0 + W,$$

где

$$H_V^0 = \begin{pmatrix} \Delta_1 & 0 & 0 & 0 & 0 \\ 0 & \Delta_2 & 0 & 0 & 0 \\ 0 & 0 & \Delta_3 & 0 & 0 \\ 0 & 0 & 0 & \Delta_4 & \Omega_{45} \\ 0 & 0 & 0 & \Omega_{45}^* & \Delta_5 \end{pmatrix}.$$

Элементы матрицы W представляют собой частоты Раби соответствующих входных пучков. Выше отмечалось, что эти пучки очень слабые, что позволяет нам рассматривать матрицу W как возмущение матрицы H_V^0 :

$$W = \begin{pmatrix} 0 & \Omega_{12} & 0 & 0 & 0 \\ \Omega_{12}^* & 0 & \Omega_{23} & 0 & 0 \\ 0 & \Omega_{23}^* & 0 & \Omega_{34} & 0 \\ 0 & 0 & \Omega_{34}^* & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Найдем собственные значения невозмущенной матрицы:

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \frac{(\Delta_4 + \Delta_5) + \sqrt{(\Delta_4 + \Delta_5)^2 - 4\Omega_{45}^2}}{2} \\ \frac{(\Delta_4 + \Delta_5) - \sqrt{(\Delta_4 + \Delta_5)^2 - 4\Omega_{45}^2}}{2} \end{pmatrix},$$

и матрицу ее собственных векторов

$$\Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{\Omega_{45}}{\sqrt{\Omega_{45}^2 + (\lambda_4 - \Delta_4)^2}} & \frac{\Omega_{45}}{\sqrt{\Omega_{45}^2 + (\lambda_5 - \Delta_4)^2}} \\ 0 & 0 & 0 & \frac{\lambda_4 - \Delta_4}{\sqrt{\Omega_{45}^2 + (\lambda_4 - \Delta_4)^2}} & \frac{\lambda_5 - \Delta_4}{\sqrt{\Omega_{45}^2 + (\lambda_5 - \Delta_4)^2}} \end{pmatrix}.$$

Поправки к собственным векторам будем искать по формуле

$$\Gamma_W = \Gamma \left(I + a^{(1)} + a^{(2)} + a^{(3)} + \dots \right),$$

где $I_{5 \times 5}$ – единичная матрица, а матрицы $a^{(i)}$ получаются из рекуррентной формулы

$$a_{m,n}^{(i)} = \frac{1}{\lambda_n - \lambda_m} \left(\left\langle \Gamma_m^{(0)} \middle| W \middle| \Gamma_n^{(0)} \right\rangle a^{(i-1)} - \sum_{k=1}^{i-1} \lambda_n^{(k)} \left\langle \Gamma_m^{(0)} \middle| \Gamma_n^{(n-k)} \right\rangle \right), \quad m \neq n.$$

Для поправок к собственным числам имеем:

$$\left\langle \Gamma_m^{(0)} \left| W \right| \Gamma_m^{(i-1)} \right\rangle = \sum_{k=1}^{i-1} E_m^k \left\langle \Gamma_m^{(0)} \left| \Gamma_m^{(i-k)} \right\rangle.$$

Нижний индекс $\Gamma_k^{(l)}$ указывает номер собственного вектора в матрице, верхний – порядок величин этого вектора.

Найдя, таким образом, собственные векторы с точностью до третьего порядка величин матрицы W , по формуле (3.1) можем посчитать значения для восприимчивости на соответствующих переходах (Ω_{ij} считаем вещественными величинами):

$$\chi_{12} = \frac{d_{12}^2}{\omega_{12}} \left(1 - \left(\frac{5 \Omega_{12}^2}{2 \omega_{12}^2} - \frac{\Omega_{23}^2}{\omega_{12} \omega_{13}} + \frac{\Omega_{12} \Omega_{23}}{\omega_{12} \omega_{13}} \right) \right),$$

$$\chi_{23} = 2 \frac{d_{23}^2 \Omega_{12}^2}{\omega_{12}^2} \left(\frac{1}{\omega_{23}} + \frac{1}{\omega_{13}} \right),$$

$$\chi_{34} = d_{34}^2 \frac{|\Omega_{12}|^2 |\Omega_{23}|^2 |\Omega_{45}|^2}{\omega_{13}^2} \times$$

$$\times \left(\frac{1}{|\Omega_{45}|^2 + (\lambda_4 - \Delta_4)^2} \left(\frac{1}{\omega_{32}^2 \omega_{34}} - \frac{1}{\omega_{12}^2 \omega_{41}} \right) + \frac{1}{|\Omega_{45}|^2 + (\lambda_5 - \Delta_4)^2} \left(\frac{1}{\omega_{32}^2 \omega_{35}} - \frac{1}{\omega_{12}^2 \omega_{51}} \right) \right)$$

В формулах для краткости введено обозначение $\omega_{ij} = \lambda_i - \lambda_j$.

Для следующих значений: $\Omega_{12} = \Omega_{23} = 0,01$; $\Omega_{45} = 0,1$; $\Delta_1 = 0$; $\Delta_2 = -4$; $\Delta_4 = -1$; $\Delta_5 = 1$; $d_{ij}/N \approx 1$; $Lk_0^i/2\pi \approx 2 \times 10^4$ ($i = \{p, c, t\}$) на рис. 3 построен график зависимости нелинейного набега фазы от величины отстройки Δ_3 .

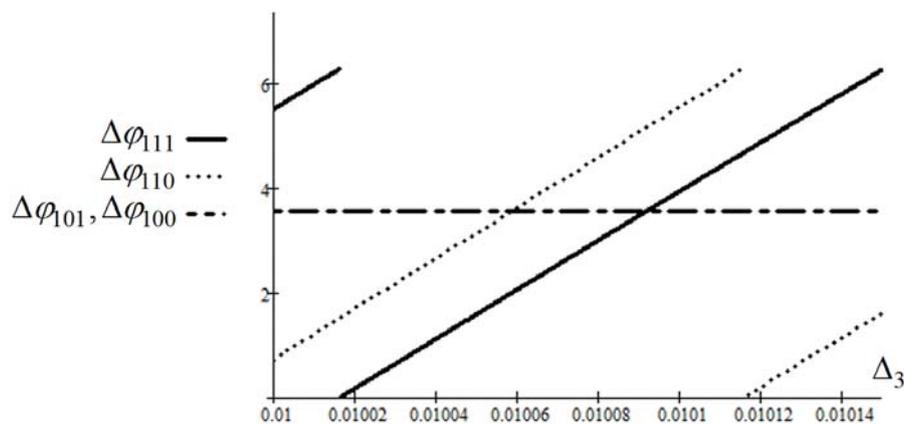


Рис. 3. Зависимость нелинейного набега фазы поляризации от величины отстройки от третьего уровня для различных круговых поляризаций входных пучков (см. табл. 1)

Из графика видно, что наибольшая разница между набегам фаз поляризации для различных случаев достигается в точке $\Delta_3 \approx 0,01006$. Для этого значения отстройки получаются следующие фазы: $\Delta\varphi_{11} \approx 0,636$ рад; $\Delta\varphi_{110} \approx 1,134$ рад; $\Delta\varphi_{101} \approx \Delta\varphi_{100} \approx 1,131$ рад. Другие значения поляризации не рассматриваются, так как в этих случаях пучки со средой не взаимодействуют, и нелинейный сдвиг фазы не наблюдается. Получить разность суммарных набегов фаз около π в случаях $\Delta\varphi_{111}$, $\Delta\varphi_{110}$, $\Delta\varphi_{101}$ и $\Delta\varphi_{100}$ мы можем контролируя Ω_{45} (рис. 4). Из графика видно, что наиболее подходящее значение $\Omega_{45} = 0,685$.

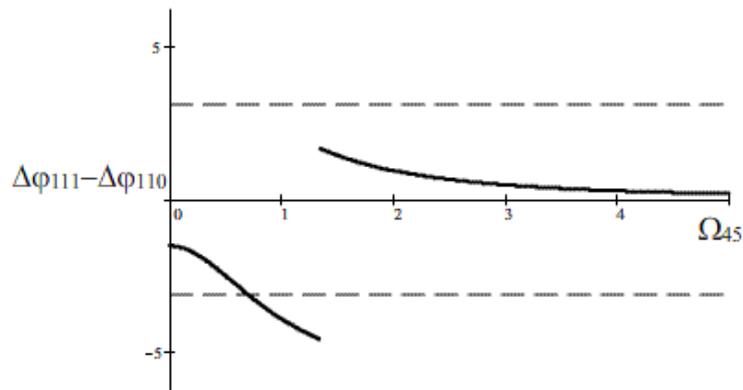


Рис. 4. Зависимость разности нелинейных набегов фаз поляризации от частоты Раби поля подстройки (рассмотрена только одна разность, так как остальные кривые близки к представленной на рисунке)

5. Заключение

Построена модель квантового логического вентиля CCNOT на основе ячейки с парами рубидия, помещенной в сильное магнитное поле. Уровни в атомах выбирались таким образом, чтобы только один вид круговой поляризации входных пучков был резонансным на данном переходе. Описано взаимодействие M-схемы со входными пучками в зависимости от их поляризации. Получены условия, при которых данная схема осуществляла контролируемый набег фазы – суммарный набег для одной из конфигураций поляризации входных пучков, существенно отличающийся от других. Приведено преобразование, позволяющее получить из контролируемого набег фазы вентиль CCNOT. Решением уравнения Шредингера получены восприимчивости на соответствующих переходах, что позволило посчитать нелинейный набег фазы поляризации для пучков в M-схеме. Наличие такого набег фазы, который существенно отличается от других, доказывает возможность построения логического вентиля.

Литература

1. Садовничий В.А. Квантовые вычисления: за и против. – Ижевск: РХД, 1999.
2. Валиев К.А., Кокин А.А. Квантовые компьютеры: надежды и реальность. – Ижевск: РХД, 2001.
3. M.D. Lukin and A. Imamoglu, Nonlinear Optics and Quantum Entanglement of Ultra-slow Single Photons. // Phys. Rev. Lett. – 2000. – 84. – 1419.
4. Lloyd S. Almost any quantum logic gate is universal // Phys. Rev. Lett. 1995. – 75. – 346.
5. C. Ottaviani et al., Polarization qubit phase gate in driven atomic media // Phys. Rev. Lett. 90, 197902 (2003).
6. C. Ottaviani et al., Quantum phase-gate operation based on nonlinear optics: full quantum analysis // Phys. Rev. A 73. – 2006. – 010301.
7. Q.A. Turchette et al., Measurement of conditional phase shift for quantum logic // Phys. Rev. Lett. – 1995. – 75. – 4710.

ПОСТРОЕНИЕ ГЕНЕРАТОРА НАЧАЛЬНОГО СОСТОЯНИЯ ДЛЯ КВАНТОВОГО КОМПЬЮТЕРА НА СВЯЗАННЫХ ЭЛЕКТРОНАХ В ВОЛНОВОДАХ

А.Е. Курасов

Научный руководитель – д.ф.-м.н., профессор И.Ю. Попов

Предложена принципиальная модель генератора начального состояния для квантового компьютера на связанных электронах в волноводах. Выведены и численно промоделированы уравнения для конкретной реализации модели с использованием одномерной ограниченной параболической потенциальной ямы.

Введение

В последнее время приобрели большую популярность исследования в области квантовых компьютеров [1, 3]. Основными направлениями этих исследований, в частности, являются различные разработки важных конструктивных частей для разнообразных реализаций квантовых компьютеров, таких как генераторы начальных состояний [8–11], процессы выполнения одно- и двухкубитных операций [4, 8–11], процессов считывания финального состояния [4, 8–11]. Данная работа посвящена разработке модели генератора начального состояния для квантового компьютера на связанных электронах в волноводах [6, 8–11].

Принципиальная модель

Начальным состоянием квантового компьютера на связанных электронах в волноводах является следующее: в каждом волноводе находится по одному электрону в нулевом состоянии, и эти электроны когерентны. Соответственно, в качестве генератора предлагается использовать устройство, состоящее из набора ловушек, по одной на каждый волновод, каждая из которых держит по одному электрону перед своим волноводом и по внешнему сигналу выпускает их в волновод достаточно быстро, чтобы эти электроны были когерентны. Т.е. необходимым требованием для ловушки является очень большое отношение времен жизни электрона в ней без сигнала и с сигналом. В качестве сигнала предлагается использовать однородное электрическое поле, которое довольно просто получить технически. В качестве самой ловушки можно взять некоторую потенциальную яму или какую-либо из моделей квантовых точек.

Вывод уравнений для частной реализации

В данной работе произведены расчеты для ловушки, состоящей из одномерной ограниченной параболической потенциальной ямы.

$$U = 0 : x < -d$$

$$U = \lambda^2 x^2 : -d \leq x \leq d ,$$

$$U = \infty : x > d$$

где d – полуширина ямы. Данный выбор формы потенциала обусловлен тем, что при добавлении к нему линейного потенциала вид уравнений останется неизменным: $\alpha x^2 + \beta x = \alpha(x + x_0)^2 + \gamma$, где x_0 пропорционально внешнему воздействию. Мы будем считать поле достаточно небольшим, чтобы можно было пренебречь вертикальным сдвигом потенциальной ямы. Время жизни электрона в яме будем рассчитывать как мнимую часть волнового числа соответствующего резонансу. Для волнового числа соответствующего резонансу было выведено следующее уравнение:

$$\theta_1 \gamma_2 + \theta_2 \gamma_1 = ik(\gamma_1 + \gamma_2), \quad (1)$$

где k – искомое волновое число, $\theta_i = \frac{\partial \psi_i(d)}{\partial x}$, $\gamma_i = \frac{\psi_i(d - x_0)}{\psi_i(d + x_0)}$, ψ_i – частные решения стационарного уравнения Шредингера в области параболического потенциала, x_0 – горизонтальный сдвиг потенциальной ямы, пропорциональный внешнему полю.

Численное моделирование

Обычно резонанс находится рядом с некоторым известным энергетическим уровнем, поэтому решение уравнения (1) искалось в виде $k = k_0 + k_1$, где k_0 соответствовало известному энергетическому уровню (использовался первый уровень бесконечной параболической потенциальной ямы). Соответственно, ψ_i представлялись в виде $\psi_i = \psi_{i0} + k_1 \psi_i^k$, где ψ_i^k находилось из исходного уравнения Шредингера путем варьирования k . Необходимым условием применимости данного приближения является $|k_1| \ll k_0$. Данным методом были рассчитаны времена жизни электрона в яме без поля, со слабым полем, с сильным полем. На рисунке изображены графики $W = \frac{\tau_0}{\tau}$, где τ_0 – время жизни электрона в яме без поля и τ – время жизни электрона в яме с полем.

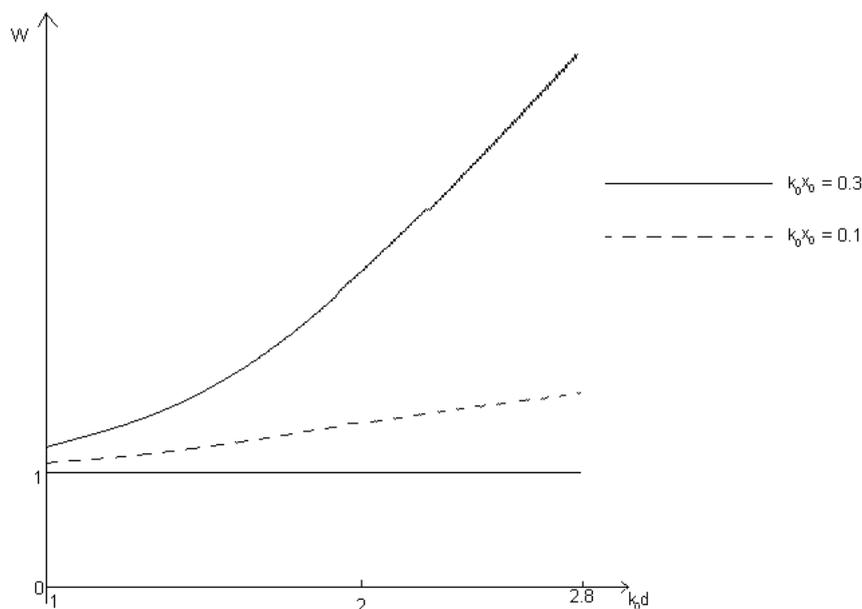


Рисунок. Отношение времен жизни электрона в яме без поля и с полем

По горизонтальной оси отложена полуширина ямы, нормированная на волновое число известного приближения. Из рисунка видно, что отношение времен растет экспоненциально с ростом ширины ямы и скорость этого роста быстро растет с ростом внешнего поля.

Вывод

Данная конструкция может быть использована в построении генератора начального состояния квантового компьютера на электронах в волноводах. Имеет смысл проанализировать альтернативные конструкции с использованием квантовых точек и потенциальных ям другого вида для получения ловушек с лучшими характеристиками.

Литература

1. Валиев К.А., Кокин А.А. Квантовые компьютеры: надежды и реальность. – М.–Ижевск: Регулярная и хаотическая динамика (РХД), 2001. – 350 с.
2. Feynman R. // *Int. J. Theor. Phys.* – 21. – 1982.
3. Feynman R. Quantum mechanical computers. // *Optics News.* – February 1985. – 11. – p. 11
4. Kane B.E. A Silicon-based Nuclear spin Quantum Computer. // *Nature.* – 1998. – V 393. – №5. – P. 133–137.
5. Tanamoto T. Quantum Computation by Coupled Quantum Dot System and Controlled NOT Operation. // 1999, LANL, E-print quant-ph/9902031, 12p
6. Wiseman H.M., Utami D.W., Sun H.B., Milburn G.J., Kane B.E., Dzurak A., Clark R.G. Quantum Measurements of Coherence in Coupled Quantum Dots //2000, LANL, E-print, arXiv:quant-ph/0002279, 19p.
7. Ландау Л.Д., Лифшиц Е.Ф. Квантовая механика. – М.: Физматгиз, 1974.
8. Gortinskaya L.V., Popov I.Yu., Tesovskaya E.S. // *Proc. of Intern. Seminar @Day on Diffraction' 2003.* St. Petersburg, 2003. P. 52.
9. Popov I.Yu., Gortinskaya L.V., Gavrilov M.I., Pestov A.A., Tesovskaya E.S. "Weakly coupled quantum wires and layers as an element of quantum computer". / *Int. Conf. "Quantum Physics and Computation"*, QPC 2005, Dubna, 2005. Abstracts p. 8.
10. Popov I.Yu., Gortinskaya L.V., Gavrilov M.I., Pestov A.A., Tesovskaya E.S. "Weakly coupled quantum wires and layers as an element of quantum computer". // *Письма в ЭЧАЯ.* – 2007. – Т.4. – №2(138). – С. 237–243.
11. Gavrilov M.I., Gortinskaya L.V., Pestov A.A., Popov I.Yu., Tesovskaya E.S. "Quantum Algorithms Implementation Using Quantum Wires System". / *Proceedings of the ICO Topical Meeting on Optoinformatics Information.*

НАШИ АВТОРЫ

- Адрианов Владимир Евгеньевич** – аспирант кафедры оптической физики и современного естествознания
- Акишина Екатерина Юрьевна** – студент кафедры оптоинформационных технологий и материалов
- Арефьева Наталья Николаевна** – научный сотрудник кафедры оптики квантоворазмерных систем
- Архипов Игорь Николаевич** – школьник лицея № 344 Невского района г. Санкт-Петербурга
- Бацын Михаил Владимирович** – аспирант Нижегородского филиала государственного университета – Высшей школы экономики
- Боголюбов Данила Александрович** – аспирант кафедры проектирования компьютерных систем
- Буданова Анна Юрьевна** – аспирант кафедры проектирования компьютерных систем
- Виноградова Галина Николаевна** – доцент кафедры оптической физики и современного естествознания
- Власов Виталий Владимирович** – аспирант кафедры проектирования компьютерных систем
- Годырева Анастасия Валерьевна** – соискатель кафедры проектирования компьютерных систем
- Голунчикова Екатерина Александровна** – аспирант Московского государственного технического университета имени Н.Э. Баумана
- Грищенко Алексей Юрьевич** – аспирант кафедры проектирования компьютерных систем
- Дрюков Николай Юрьевич** – аспирант кафедры проектирования компьютерных систем
- Елисеев Олег Валерьевич** – аспирант кафедры проектирования компьютерных систем
- Ермаков Иван Владимирович** – аспирант кафедры проектирования компьютерных систем
- Ермаков Николай Владимирович** – аспирант кафедры проектирования компьютерных систем
- Жуков Сергей Николаевич** – студент кафедры оптоинформационных технологий и материалов
- Замотин Кирилл Юрьевич** – аспирант Санкт-Петербургского государственного политехнического университета
- Зимин Владимир Николаевич** – аспирант кафедры проектирования компьютерных систем
- Иванова Светлана Михайловна** – студент кафедры проектирования компьютерных систем
- Керпелева Светлана Юрьевна** – сотрудник кафедры нанотехнологий и материаловедения
- Клементьева Анастасия Валерьевна** – студент кафедры оптоинформационных технологий и материалов
- Клоков Михаил Викторович** – студент кафедры нанотехнологий и материаловедения
- Козак Владимир Александрович** – аспирант кафедры проектирования компьютерных систем
- Козлицкая Анна Васильевна** – студент Вятского государственного университета
- Козлов Арсений Анатольевич** – студент Пермского государственного университета

Кощеев Алексей Васильевич – аспирант Вятского государственного университета
Кривошеев Александр Анатольевич – студент кафедры проектирования компьютерных систем
Кручинин Станислав Юрьевич – аспирант кафедры оптической физики и современного естествознания
Крылов Вадим Анатольевич – аспирант кафедры проектирования компьютерных систем
Кувшинов Станислав Сергеевич – аспирант кафедры проектирования компьютерных систем
Курасов Александр Евгеньевич – аспирант кафедры высшей математики
Лазарева Ксения Евгеньевна – студент кафедры оптоинформационных технологий и материалов
Левицкий Руслан Сергеевич – научный сотрудник Центра «Оптические информационные технологии»
Лесничий Василий Валерьевич – студент кафедры фотоники и оптоинформатики
Липатов Алексей Леонидович – аспирант кафедры проектирования компьютерных систем
Логунов Александр Евгеньевич – аспирант центра «Информационные оптические технологии»
Лукашук Вероника Олеговна – ассистент Уфимского государственного авиационного технического университета
Лукашук Станислав Юрьевич – доцент Уфимского государственного авиационного технического университета
Лукин Михаил Александрович – студент Вятского государственного университета
Миноженко Ольга Александровна – студент кафедры компьютерной фотоники
Михайленко Алексей Станиславович – аспирант кафедры проектирования компьютерных систем
Михайличенко Ольга Викторовна – соискатель кафедры проектирования компьютерных систем
Москаленко Станислав Владимирович – аспирант кафедры проектирования компьютерных систем
Нечаев Виталий Александрович – соискатель кафедры проектирования компьютерных систем
Нечаева Наталья Викторовна – ассистент кафедры проектирования компьютерных систем
Николаева Татьяна Сергеевна – аспирант кафедры проектирования компьютерных систем
Орлова Анна Олеговна – научный сотрудник центра «Информационные оптические технологии»
Папко Александр Вадимович – соискатель кафедры лазерной техники и биомедицинской оптики
Парфенов Петр Сергеевич – сотрудник кафедры оптической физики и современного естествознания
Пирожникова Ольга Игоревна – студент кафедры проектирования компьютерных систем
Прохожев Николай Николаевич – аспирант кафедры проектирования компьютерных систем
Пудов Денис Владимирович – аспирант кафедры проектирования компьютерных систем
Свитенков Андрей Игоревич – студент кафедры фотоники и оптоинформатики

Семёнов Вениамин Александрович – аспирант кафедры проектирования компьютерных систем

Семёнова Мария Александровна – аспирант кафедры проектирования компьютерных систем

Семьина Светлана Алексеевна – студент кафедры компьютерной фотоники

Ситдигов Владимир Масгутович – студент кафедры оптоинформационных технологий и материалов

Солнушкин Константин Сергеевич – аспирант Санкт-Петербургского государственного политехнического университета

Соловьёв Виктор Павлович – аспирант кафедры проектирования компьютерных систем

Соловьёв Денис Викторович – аспирант кафедры проектирования компьютерных систем

Старовойтов Антон Андреевич – аспирант кафедры оптической физики и современного естествознания

Степанов Антон Андреевич – сотрудник центра «Информационные оптические технологии»

Стройков Илья Игоревич – аспирант кафедры проектирования компьютерных систем

Студеникин Олег Леонидович – сотрудник кафедры проектирования компьютерных систем

Трифанов Александр Игоревич – студент кафедры высшей математики

Туранцев Дмитрий Сергеевич – аспирант кафедры проектирования компьютерных систем

Увашев Руслан Сергеевич – студент Вятского государственного университета

Уланов Александр Ильич – аспирант Санкт-Петербургского государственного политехнического университета

Ушакова Елена Владимировна – студент кафедры оптической физики и современного естествознания

Федосов Павел Валентинович – аспирант кафедры проектирования компьютерных систем

Федотов Андрей Сергеевич – аспирант кафедры проектирования компьютерных систем

Фетинина Анна Игоревна – студент Вятского государственного университета

Французов Григорий Сергеевич – студент кафедры нанотехнологий и материаловедения

Цингер Елена Олеговна – аспирант кафедры проектирования компьютерных систем

Цыганкова Екатерина Владимировна – студент кафедры оптоинформационных технологий и материалов

Черевков Сергей Александрович – студент кафедры лазерной техники и биомедицинской оптики

Шилкин Денис Андреевич – аспирант кафедры проектирования компьютерных систем

Юдин Дмитрий Геннадьевич – аспирант кафедры проектирования компьютерных систем

Научная школа «ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ, ПРОЕКТИРОВАНИЕ, ТЕХНОЛОГИИ ЭЛЕМЕНТОВ И УЗЛОВ КОМПЬЮТЕРНЫХ СИСТЕМ»		3
СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ.....		4
Боголюбов Д.А. Применение элементов теории графов в конечно-элементном анализе		4
Федотов А.С. Анализ автоматизированного рабочего места как объекта проектирования		8
Москаленко С.В. Волновой алгоритм векторизации линейных растровых изображений		16
Власов В.В. Возможности граф-ориентированного программирования для систем автоматизированного управления процессами на примере jPDL языка		22
Соловьёв Д.В. Нейросетевой метод оптимизации математических моделей сложных технологических процессов		33
Федосов П.В. Использование САПР при создании «умных домов».....		39
Шилкин Д.А. Элементы оптической подсистемы СКС		44
МИКРОЭЛЕКТРОНИКА, ДЕФЕКТОСКОПИЯ И ДЕФЕКТООБРАЗОВАНИЕ В ПРОЦЕССАХ ПРОИЗВОДСТВА И ЭКСПЛУАТАЦИИ ЭЛЕМЕНТНОЙ БАЗЫ И СУ		55
Стройков И.И. Исследование нанокompозитных структур на основе кремния и фосфатного стекла на кремниевых подложках		55
Буданова А.Ю., Крылов В.А. Анализ современной патентной литературы по микромеханическим акселерометрам		63
Нечаев В.А., Нечаева Н.В., Пирожникова О.И. Бесконтактный метод контроля механических параметров ферромагнитных материалов		71
Нечаев В.А., Нечаева Н.В., Пирожникова О.И. Магнитный метод неразрушающего контроля механических характеристик широкополосных сталей.....		73
Нечаев В.А., Нечаева Н.В., Пирожникова О.И. Исследование влияния динамических нагрузок на величину деформации верхнего слоя ферромагнетиков		75
Иванова С.М., Елисеев О.В., Михайленко А.С., Студеникин О.Л. Обзор методов первичных параметров оптического волокна.....		78
БИОТЕХНИЧЕСКИЕ ИНФОРМАЦИОННО-ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ		84
Грищенко А.Ю. Уточнение рабочего частотного диапазона аппаратно-программного комплекса СТАП.....		84
ПЕРСПЕКТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ.....		88
Цингер Е.О. Формирование структурно-функциональной связи «заказчик-пользовательская сеть» при создании СУ КМП в системе ОМС		88
Юдин Д.Г. Поддержание качества Web-приложения, написанного на языке PHP		93
Пудов Д.В. Обзор решений для управления ИТ-инфраструктурой		99

Зимин В.Н., Федосов П.В. Разработка ВЕБ-приложений на основе технологии AJAX в среде Microsoft Visual Studio	103
Николаева Т.С. Представление информационных потоков с помощью теории графов	108
Козак В.А. Разработка автоматизированной системы обслуживания абонентов сети IP-телефонии на базе протокола SIP	112
Соловьёв В.П. Система контроля за системой через лог-файлы	122
Архипов И.Н. Разработка и изготовление электронного устройства для проведения опросов	129
МЕТОДЫ ЗАЩИТЫ ИНФОРМАЦИИ.....	133
Голунчикова Е.А., Липатов А.Л. Методика оценки защищенности конфиденциальной информации, обрабатываемой основными техническими средствами, от ее утечки за счет побочных электромагнитных излучений, при использовании средств активной защиты	133
Туранцев Д.С. Проблемы безопасности в беспроводных ЛВС IEEE 802.11	136
Годырева А.В. Возможные каналы утечки информации на предприятии.....	146
Кувшинов С.С. Система скрытой передачи сообщений в графических изображениях формата JPEG	152
Семёнова М.А., Семёнов В.А. Комплексные решения в области соблюдения мер обеспечения информационной безопасности в банковской сфере	160
Михайличенко О.В., Прохожев Н.Н. Оценка устойчивости ЦВЗ к внешним воздействиям, внедренных с помощью алгоритмов пространственной области встраивания.....	168
Кривошеев А.А. Информационная война. Информационное оружие. Современная ситуация.....	173
Дрюков Н.Ю., Ермаков И.В., Ермаков Н.В. Методика построения модели угрозы информационной безопасности	181
Научная школа «НАНОМАТЕРИАЛЫ, НАНОСТРУКТУРЫ, НАНОТЕХНОЛОГИИ»	186
Степанов А.А., Орлова А.О. Спектрально-люминесцентное исследование взаимодействия квантовых точек с тетрапиррольными молекулами.....	187
Левицкий Р.С. Эффект фотонной лавины в наноструктурах	196
Семьина С.А. Исследование структуры и свойств оптических УФ-отверждаемых акрилатных полимеров, наполненных наночастицами ZnO.....	207
Миноженко О.А. Концентрационное тушение люминесценции европия в органо-неорганическом композите	213
Керпелева С.Ю., Клоков М.В., Французов Г.С. Визуализация микро- и наноструктуры клеток в нативном состоянии с измерением жесткости и электрической проводимости клеточных мембран	218
Арефьева Н.Н. Наноимпринт – формирование нано- и микро- элементов фотоники контактным копированием	229
Кручинин С.Ю. Резонансный перенос энергии фотовозбуждения в квантовых точках	235
Лазарева К.Е., Акишина Е.Ю. Влияние Вг на фото-термо-индуцированную кристаллизацию стекла.....	245
Акишина Е.Ю., Лазарева К.Е. Оптические свойства фото-термо-рефрактивных стекол с различным содержанием фтора	251

Ситдиков В.М., Жуков С.Н. Поляризованная люминесценция в стеклах и наностеклокерамике, активированных ионами переходного четырехвалентного хрома	257
Клементьева А.В. Спектрально-люминесцентные свойства прозрачной свинцово-фторидной наностеклокерамики, активированной ионами эрбия	264
Цыганкова Е.В. Влияние сурьмы на формирование наночастиц серебра в фото-термо-рефрактивном стекле	272
Парфенов П.С. Реализация одноканального счета фотонов для рамановского наноспектрометра.....	278
Черевков С.А. Техника микро-рамановского анализа материалов	286
Ушакова Е.В. Перенос энергии фотовозбуждения в системах квантовых точек	294
Адрианов В.Е., Виноградова Г.Н. Самоорганизация квантовых нанокристаллов CdTe на гидрофобной поверхности.....	300
Свитенков А.И., Лесничий В.В. Элементарные течения в наногидродинамике	309
Логунов А.Е., Папко А.В. Оптические методы определения размеров и форм наночастиц	316
Старовойтов А.А. Перестройка структуры наноконпонентов молекулярного слоя под действием нагрева и оптического излучения	322

**Научная школа «ТЕХНОЛОГИИ
ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ
ВЫЧИСЛЕНИЙ И СИСТЕМ»**

Уланов А.И. Математическое моделирование геомеханических процессов.....	331
Замотин К.Ю. Расчет акустических характеристик автомобильных зеркал	338
Лукин М.А., Увашев Р.С. Анализ потока изображений на основе искусственных нейронных сетей в реальном масштабе времени	348
Козлицкая А.В., Кошеев А.В., Фетинина А.И. Идентификация параметров математической модели развивающейся экономики	354
Бацын М.В. Об одном алгоритме вычисления функции распределения выплат в модели коллективных страховых рисков	361
Козлов А.А. Мультиагентная система динамической балансировки распределенной имитационной модели	373
Лукащук В.О., Лукащук С.Ю. Образовательная программа краткосрочного повышения квалификации в области суперкомпьютерных технологий	
Солнушкин К.С. Проблемы анализа и моделирования производительности ЭВМ....	383
Трифанов А.И. Реализация квантового логического вентиля CCNOT на основе зеемановской структуры уровней в атомах рубидия	390
Курасов А.Е. Построение генератора начального состояния для квантового компьютера на связанных электронах в волноводах.....	399

НАШИ АВТОРЫ..... 402

Научно-технический вестник СПбГУ ИТМО. Выпуск 51. НАУЧНЫЕ ШКОЛЫ В СПбГУ ИТМО / Главный редактор д.т.н., проф. В.О. Никифоров. – СПб: СПбГУ ИТМО, 2008. – 408 с.

**НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК СПбГУ ИТМО
Выпуск 51**

НАУЧНЫЕ ШКОЛЫ В СПбГУ ИТМО

Главный редактор
доктор технических наук, профессор
В.О. Никифоров
Дизайн обложки В.А. Петров, А.А. Колокольников
Редакционно-издательский отдел СПбГУ ИТМО
Зав. РИО Н.Ф. Гусарова
Лицензия ИД № 00408 от 05.11.99.
Подписано в печать 10.04.08.
Заказ 1191. Тираж 100 экз.